

# **NATIONAL SURVEY ON DRUG USE AND HEALTH: SUMMARY OF METHODOLOGICAL STUDIES, 1971–2014**

## **DISCLAIMER**

SAMHSA provides links to other Internet sites as a service to its users and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.

Substance Abuse and Mental Health Services Administration  
Center for Behavioral Health Statistics and Quality  
Rockville, Maryland 20857

November 2014

*This page intentionally left blank*

# **NATIONAL SURVEY ON DRUG USE AND HEALTH: SUMMARY OF METHODOLOGICAL STUDIES, 1971–2014**

Contract No. HHSS283201000003C  
RTI Project No. 0212800.001.208.007.034

RTI Authors:

Hyunjoo Park  
Doug Currivan  
Kevin Wang

SAMHSA Authors:

Sarra Hedden  
Arthur Hughes  
Dicy Painter

RTI Project Director:

David Hunter

SAMHSA Project Officer:

Peter Tice

For questions about this report, please email [Peter.Tice@samhsa.hhs.gov](mailto:Peter.Tice@samhsa.hhs.gov).

Prepared for Substance Abuse and Mental Health Services Administration,  
Rockville, Maryland

Prepared by RTI International, Research Triangle Park, North Carolina

November 2014

Recommended Citation: Center for Behavioral Health Statistics and Quality.  
(2014). *National Survey on Drug Use and Health (NSDUH): Summary of  
Methodological Studies, 1971–2014*. Substance Abuse and Mental Health  
Services Administration, Rockville, MD.

## **Acknowledgments**

This document was originally prepared and published in 2006 by the Office of Applied Studies (OAS, the former name for the Center for Behavioral Health Statistics and Quality [CBHSQ]), Substance Abuse and Mental Health Services Administration (SAMHSA), and by RTI International (a registered trademark and a trade name of Research Triangle Institute), Research Triangle Park, North Carolina. In 2006, work by RTI was performed under Contract No. 283-03-9028. Contributors and reviewers at RTI listed alphabetically included Tim Flanigan, Mary Ellen Marsden, Emily McFarlane, Joe Murphy, Hyunjoo Park, Lanny L. Piper, Adam Safir, Thomas G. Virag (Project Director), and Kevin Wang. Contributors at SAMHSA listed alphabetically included Joseph Gfroerer and Arthur Hughes. At RTI, Richard S. Straw edited the report with assistance from Jason Guder and Claudia M. Clark. Also at RTI, Roxanne Snaauw formatted and word processed the report, and Pamela Couch Prevatt, Teresa F. Gurley, and Shari B. Lambert prepared its press and Web versions. Final report production was provided by Beatrice Rouse, Coleen Sanderson, and Jane Feldman at SAMHSA.

For this 2014 updated edition of the document, contributors and reviewers at RTI listed alphabetically include James R. Chromy, Jennifer Cooney, Devon S. Cribb, Doug Currivan, Rebecca A. Granger, Dave Hunter (Project Director), Larry A. Kroutil, Lisa E. Packer, Hyunjoo Park, and Kevin Wang. Contributors at SAMHSA listed alphabetically include Sarra Hedden, Art Hughes, Dicy Painter, and Peter Tice (Project Officer). At RTI, Richard S. Straw edited the report. Also at RTI, Debbie F. Bond and Judy B. Cannada formatted and word processed the report, and Teresa F. Bass, Kimberly H. Cone, Pamela Couch Prevatt, and Pamela Tuck prepared its Web version.

## Table of Contents

Chapter	Page
Introduction .....	1
Purpose of Report .....	1
Structure .....	1
Sources .....	2
Survey Background and History of Methodological Studies .....	4
1975–1986 .....	7
1990–1991 .....	13
1992 .....	17
1993 .....	29
1994 .....	33
1995 .....	39
1996 .....	41
1997 .....	49
1998 .....	57
1999 .....	67
2000 .....	71
2001 .....	77
2002 .....	93
2003 .....	105
2004 .....	113
2005 .....	123
2006 .....	143
2007 .....	147
2008 .....	155
2009 .....	161

## Table of Contents (continued)

<b>Chapter</b>	<b>Page</b>
2010 .....	163
2011 .....	177
2012 .....	183
2013 .....	193
2014 .....	209
Author and Subject Index .....	219

# Introduction

## Purpose of Report

This report provides a comprehensive summary of methodological research conducted on the National Survey on Drug Use and Health (NSDUH). Since its inception in 1971, considerable research has been conducted on the methods used in the survey. Studies have addressed sampling, counting and listing operations, questionnaire design, data collection procedures, editing and imputation procedures, nonresponse, and statistical analysis techniques associated with the survey. This research has been critical to the NSDUH project, providing information to identify and quantify survey errors, to guide redesign efforts, and to develop more efficient survey processes. These research studies have been published in professional journals, books and book chapters, conference proceedings of professional association meetings, and reports published by Federal agencies, such as the Substance Abuse and Mental Health Services Administration (SAMHSA) and the National Institute on Drug Abuse (NIDA).

The summaries of published studies included in this report provide a record of the wealth of methodological findings stemming from NSDUH research. This research is of great value to users of NSDUH data who need to understand the impact of method effects on their own studies. Researchers using NSDUH data for epidemiological studies need to understand these effects when designing their analysis plans and also should be aware of these issues in interpreting their results. The NSDUH body of methodological research also will be useful to those responsible for managing large-scale surveys or designing new surveys, particularly if they will collect substance use data. Compiling all of these NSDUH-related findings in one document will make it much easier for these researchers to find the studies most relevant for their purposes.

## Structure

The report is organized into chapters based on year of publication. Within each chapter, publication summaries are sorted alphabetically by the last name of the first author. For increased readability, publication summaries are further organized into four uniform sections. The Citation section provides bibliographic detail. The Purpose/Overview section presents the motivation for the research and reviews the history or background of the subject matter, if available. For experiments or other analyses, the Methods section describes the methods used to conduct the research; for reports covering multiple method initiatives, or that describe or summarize the results of other studies, the Methods section presents an overview of the organization of the publication. The Results/Conclusions section summarizes the findings. When the nature of the publication is such that there are no methods or results/conclusions per se (e.g., introductory comments to an invited paper session published in a proceedings volume), an "N/A" appears in the section.

For publications in the public domain, such as those prepared by RTI International<sup>1</sup> (the NSDUH contractor since 1988) or SAMHSA, the original verbatim abstract was used as a starting point for the final edited summary. Summaries of copyrighted publications, such as

---

<sup>1</sup> RTI International is a registered trademark and a trade name of Research Triangle Institute.

books or journal articles, were written independently of existing abstracts after substantial review of the source material. In all cases, the summaries in this report present mainly key points from source publications. Therefore, readers are encouraged to access the source publication for complete information.

## Sources

The publications summarized in this report were compiled through an extensive search of archives and databases. The following keywords were used during the search of the literature:

- NSDUH,
- National Survey(s) on Drug Use and Health,
- NHSDA,
- National Household Survey(s) on Drug Abuse,
- National Analytical Center,
- NAC and drug(s) and survey(s),
- NSDUH Methodological Procedures,
- NHSDA Methodological Procedures,
- NSDUH Methodology and Research Methods, and
- NHSDA Methodology and Research Methods.

The vast majority of the records found in most of the sources consisted only of secondary data analyses, results of the study, or data used without mentioning any methodological issues. Records that were methodology related were identified through titles, abstracts, and keywords. When unsure of whether or not a record was methodology related, the full article was reviewed when available. A list of the sources used to prepare the report is provided below. The source publications for much of the research summarized in this report are available online and can be found by visiting the Web sites listed in the Archives and Databases section below.

### *Archives and Databases*

- Conference Programs of the American Association for Public Opinion Research (AAPOR):  
[http://www.aapor.org/default.asp?page=conference\\_and\\_events/past\\_conferences](http://www.aapor.org/default.asp?page=conference_and_events/past_conferences)
- Proceedings of the Survey Research Methods Section of the American Statistical Association (ASA):  
<http://www.amstat.org/sections/srms/Proceedings/>

- Catalog of U.S. Government Publications (CGP):  
<http://catalog.gpo.gov/F>
- Proceedings of the Conference on Health Survey Research Methods (CHSRM):  
<http://www.cdc.gov/nchs/products/hsrmc.htm>
- Entrez, the Life Sciences Search Engine (National Center for Biotechnology Information):  
<http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>
- Education Resource Information Center (ERIC):  
<http://www.eric.ed.gov/ERICWebPortal/Home.portal>
- Government Accountability Office (GAO):  
<http://www.gao.gov/>
- Scholarly Journal Archive (JSTOR):  
<http://www.jstor.org/>
- MathSciNet:  
<http://www.ams.org/mathscinet>
- *Medical Care*, Official Journal of the Medical Care Section, American Public Health Association:  
<http://www.lww-medicalcare.com/pt/re/medcare/home.htm>
- National Academies Press:  
<http://www.nap.edu/>
- National Clearinghouse for Alcohol and Drug Information (NCADI), now the SAMHSA Store:  
<http://store.samhsa.gov/home>
- National Criminal Justice Reference Service (NCJRS) Abstracts Database:  
<http://www.ncjrs.gov/abstractdb/search.asp>
- National Institute on Drug Abuse (NIDA):  
<http://www.nida.nih.gov/>
- National Survey on Drug Use and Health (NSDUH):  
<http://www.samhsa.gov/data/population-data-nsduh>
- PsycINFO:  
<http://www.apa.org/psycinfo/>
- PubMed (National Library of Medicine):  
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>

- Substance Abuse and Mental Health Data Archive (SAMHDA):  
<http://www.icpsr.umich.edu/SAMHDA/>
- Sociological Abstracts:  
<http://www.proquest.com/products-services/socioabs-set-c.html>
- Web of Science (includes Arts and Humanities Citation Index, Science Citation Index Expanded, and Social Sciences Citation Index):  
<http://thomsonreuters.com/thomson-reuters-web-of-science/>
- WorldCat (combined book catalog of over 30,000 libraries):  
<http://www.oclc.org/worldcat/default.htm>

### *Methodological Reports*

- Survey Measurement of Drug Use: Methodological Studies, edited by Turner, Lessler, and Gfroerer;
- Redesigning an Ongoing National Household Survey: Methodological Issues, edited by Gfroerer, Eyerman, and Chromy; and
- Evaluating and Improving Methods Used in NSDUH, edited by Kennet and Gfroerer.

In addition to these archives, databases, and methodological reports, internal RTI lists of conference papers, dissertation abstracts, and staff resumes were reviewed for NSDUH methodological citations. Internet search engines, such as Google.com and iTools.com, also were used to find citations of interest.

### **Survey Background and History of Methodological Studies**

NSDUH is an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years old or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA). It is the primary source of statistical information on the use of illegal drugs, alcohol, and tobacco by the U.S. civilian, noninstitutionalized population aged 12 or older. The survey also includes several modules of questions that focus on mental health issues. Conducted by the Federal Government since 1971, the survey collects data by administering questionnaires to a representative sample of the population through face-to-face interviews at their places of residence. NSDUH is sponsored by SAMHSA, U.S. Department of Health and Human Services, and is planned and managed by SAMHSA's Center for Behavioral Health Statistics and Quality (CBHSQ), formerly the Office of Applied Studies (OAS). Data collection and analysis are conducted under contract with RTI International, Research Triangle Park, North Carolina.

The first two surveys in 1971 and 1972 were done under the auspices of the National Commission on Marihuana and Drug Abuse. These initial surveys of about 3,000 respondents each year employed data collection procedures designed to elicit truthful responses to questions on drug use. The methodology included the use of anonymous, self-administered answer sheets

completed in a private place within respondents' homes. NHSDA was continued on a periodic basis by NIDA, starting in 1974, using the same basic design. As the demand for data and the importance of the survey increased, the sample size was increased, and the questionnaire and methodology were refined. In particular, the increasing impact of cocaine use in the United States in the 1980s led to significant increases in the sample, driven in part by the needs of the newly created Office of National Drug Control Policy, which required improved data for tracking illicit drug use. With the larger survey cost and more prominent role of NHSDA data among policymakers, the need for more research on the NHSDA methods was clear. Thus, beginning in the late 1980s, methodological research became an ongoing component of the NHSDA project under NIDA. When the project was transferred to OAS in the newly created SAMHSA in 1992, the methodology program continued.

The initial series of studies (most were published in Turner, Lessler, & Gfroerer, 1992) evaluated the basic survey methods and tested possible improvements, ultimately leading to a redesign of the questionnaire in 1994, using the same basic data collection method (self-administered paper answer sheets). Soon after this redesign, however, SAMHSA decided in 1995 to pursue the use of a newly emerging data collection technology called audio computer-assisted self-interviewing (ACASI). Thus, NHSDA methodological research turned to the testing and development of ACASI, which SAMHSA ultimately implemented in 1999 (OAS, 2001). The ACASI implementation occurred simultaneously with a new sampling plan, designed to produce State-level estimates, with a sample size of about 70,000 respondents per year. The implementation of the 1999 redesign generated a number of methodological results, stemming both from the development of the new design and from the assessment of its impact (Gfroerer, Eyerman, & Chromy, 2002). Other interesting methodological findings emerged following the 1999 redesign; in 2002 and beyond, with the testing and implementation of respondent incentives and when the name of the survey was changed to the National Survey on Drug Use and Health; and with other assessments of the survey procedures (Kennet & Gfroerer, 2005).

As noted in 2006 when it was first published, this report will be updated on a regular basis to include summaries for new methodological publications. Although every effort has been made to include citations for all important methodological studies published to date, it is possible that some have been omitted. Readers are encouraged to notify CBHSQ about any omissions by visiting the following Web page: <http://www.samhsa.gov/about-us/contact-us>.

For inquiries on the availability of specific NSDUH research and other questions, visit <http://www.samhsa.gov/data/population-data-nsduh>. To order printed copies of available reports, visit <http://store.samhsa.gov/home>.

*This page intentionally left blank*

## 1975–1986

### **The effectiveness of cash incentives to respondents as an inducement to cooperate in completing an in-home interview: Fall 1981 incentive evaluation report for the National Drug Abuse Survey**

CITATION: Abelson, H. I., & Fishburne, P. M. (1981, December). *The effectiveness of cash incentives to respondents as an inducement to cooperate in completing an in-home interview: Fall 1981 incentive evaluation report for the National Drug Abuse Survey* (RAC 4305, prepared for the Social Research Group at George Washington University and the National Institute on Drug Abuse). Rockville, MD: National Institute on Drug Abuse.

PURPOSE/OVERVIEW: In the late 1970s, survey organizations reported increasing difficulty in obtaining respondent cooperation. A growing body of literature suggested that one means of reducing respondent noncooperation was to use an incentive (i.e., a reimbursement to respondents for the burden of responding). Such incentives were seen as a means of enhancing respondent motivation, reducing respondents' perceptions of inconvenience, reducing item nonresponse, and improving interviewers' expectations of respondent cooperation. In surveys prior to the 1982 National Household Survey on Drug Abuse (NHSDA), respondents received a gratuity valued at approximately \$3.00. This practice was continued in the 1982 survey. In an effort to determine the cost-effectiveness of this practice in reducing interviewer callbacks and ensuring at least an 80 percent response rate, an incentive evaluation study was undertaken as part of the 1982 survey.

METHODS: During the initial 30-day field data collection period, the contractor conducted an evaluation of the incentive given to the respondents. A cluster of approximately 300 households was assigned to different conditions. The control conditions were defined as no incentives given to respondents. Two experimental conditions (approximately 100 respondents each) were defined in terms of the size of the incentive. Each of the experimental groups received \$3.00 and \$6.00, respectively. The authors performed an empirical test of the effectiveness and cost-efficiency of incentives. Specifically, the authors compared the use of \$3.00 and \$6.00 incentives with the use of no incentive to test the hypotheses that (1) the use of incentives increases the response rate, (2) the use of incentives decreases the average number of visits to the households, and (3) the use of incentives decreases the average direct cost per completed interview.

RESULTS/CONCLUSIONS: The data showed no statistically significant differences attributable to the incentive treatments. Specifically, with regard to the three hypotheses tested, there were no statistically significant differences between the three treatment levels in terms of response rate, average number of visits to the household, or average direct cost per completed interview.

## Sensitivity of survey techniques in measuring illicit drug use

**CITATION:** Cisin, I. H., & Parry, H. J. (1975, March). *Sensitivity of survey techniques in measuring illicit drug use* (Report No. 75-1, Contract No. HSM 42-73-197 (ND)). Rockville, MD: National Institute on Drug Abuse.

**PURPOSE/OVERVIEW:** By the mid-1970s, the problem of validity of response was one that had troubled sample survey researchers ever since the great expansion of the field in the 1930s. A considerable number of methodological studies, many of the most ingenious kind, had addressed themselves to the query: "How much credence can we give to behavior reported by respondents, particularly when self-esteem or deviant behavior is involved?" Most studies of validity emphasized the expectation that social pressure on respondents may result in the underreporting of behavior that could be called deviant or at least inappropriate and the overreporting of behavior that can add to self-esteem. Sometimes, of course, researchers for one reason or another may miscalculate the amount and direction of social pressure to which their respondents are subjected. Kinsey, for example, in his pioneering study of male sexuality was laudably desirous of steering at safe distance from the Scylla of underreporting certain types of sexual behavior. In his attempt to avoid an obvious danger, in the view of some critics, he passed too close to the Charybdis of overreporting.

Methodological studies in the field of validity of response had, in general, been of two kinds by the mid-1970s. The first type consisted of a comparison of two groups: an index group that because of known or assumed past behavior was significantly more likely than a control group to have acted in a certain way. A survey conducted among matched samples of the two groups should yield results showing significantly greater amounts of the behavior being studied among the index group. The second type of validity study was where reported behavior of a "known" group is compared with actual previously recorded behavior. Prior to the full-scale national survey on drug abuse being conducted in 1975, it was decided to precede the survey with a special validity study to examine the responses of a "known" index group with those of a control group. Additionally, after the validity study went into the field, the authors decided to augment it with a comparison of survey responses with records.

**METHODS:** The index group consisted of 100 respondents who had sought and received treatments at certain drug abuse clinics in the San Francisco Bay area. The control group consisted of 100 randomly selected respondents who were matched against a member of the index group in terms of age, gender, and location, but who had not sought help at a drug abuse clinic. Each group was asked a series of questions concerning a large number of drugs, questions dealing with past use and current use, with the influence of the peer group, and with past year incidence (i.e., use of a drug for the first time in the 12 months preceding the survey). Responses of the two groups then were compared, both in terms of individual drug behavior and in use of various combinations of drugs.

**RESULTS/CONCLUSIONS:** The study found a large number of significant differences in drug behavior between the two groups. Such differences tended to be concentrated at the beginning of the drug use continuum (i.e., the "used at some time" column). Much less difference was found in the "current use" percentage partly because current use for any given drug (except marijuana)

was generally low, so that differences were not statistically significant. Nevertheless, in nearly all cases the differences were in the expected direction.

### **Heroin incidence: A trend comparison between national household survey data and indicator data**

**CITATION:** Crider, R. A. (1985). Heroin incidence: A trend comparison between national household survey data and indicator data. In B. A. Rouse, N. J. Kozel, & L. G. Richards (Eds.), *Self-report methods of estimating drug use: Meeting current challenges to validity* (pp. 125-140). Rockville, MD: National Institute on Drug Abuse.

**PURPOSE/OVERVIEW:** A commonly held belief in the mid-1980s was that respondents in a face-to-face survey will underreport the extent of their involvement in deviant behavior (Fishburn, 1980; Miller, 1985 [see abstract in this volume]; Sirkin, 1975). To test this assumption, a time series of self-reported year of first use of heroin was compared with a time series of heroin indicator data, such as heroin-related emergencies and deaths, treatment admissions, and hepatitis B cases. Although there was a general belief that self-reported heroin use would be underreported, no test of that assumption had been conducted comparing the time-series trend of heroin incidence based on survey data to the time-series trends of heroin indicators.

**METHODS:** Seven drug indicators were compared: (1) the number of heroin initiates based on the face-to-face National Household Survey on Drug Abuse (National Survey); (2) the number of heroin initiates voluntarily entering a panel of consistently reporting federally funded treatment programs (Client-Oriented Data Acquisition Process, CODAP) for the first time by the year of first use; (3) the residual number of hepatitis B cases per year; (4) the percentage of high school seniors ever using heroin (Johnston et al., 1982); (5) the number of heroin-related emergency room visits reported to the Drug Abuse Warning Network (DAWN); (6) the number of heroin-related deaths reported to DAWN; and (7) the average street-level heroin purity (U.S. Drug Enforcement Administration [DEA], 1984). Because of the small proportion of the population reporting ever having used heroin, the year of first use data from the National Institute on Drug Abuse's (NIDA's) National Surveys conducted in 1977, 1979, and 1982 were pooled to show the number of new users in the household population by year of first use. In addition, the data were "smoothed" by using a 2-year moving average.

**RESULTS/CONCLUSIONS:** The number of cases of heroin use combined from the three National Surveys were shown by year of first use, 1965 through 1980. Although the number of heroin users in any one survey was small, the pooled data produced frequencies large enough to establish trends. These data were not used to make estimates of the number of heroin initiates in any particular year, but were used to show a changing pattern over a several-year period. The epidemics in the early 1970s and the mid-1970s were evident. These epidemic periods occurred at the same periods reported by high school seniors and by heroin users in treatment. The household self-report data trends based on age and frequency of use also were consistent with the trends in periods of initiation reported by heroin users in treatment as noted in drug abuse treatment admission data for year of first heroin use. Trends in indicators of heroin epidemics were compared with trends based on self-report data from the National Surveys. The trends in

hepatitis B cases, heroin-related emergency room visits, heroin-related deaths, and the average retail heroin purity were consistent with the epidemic periods suggested by the household data. This consistency among the three sources of self-reported data on trends in year of first heroin use, combined with the consistency of these self-reported data with the trends based on the indicators of heroin epidemics, offers some validation to the use of retrospective direct questions concerning age of first use of heroin to monitor heroin incidence in the household population.

## **Influence of privacy on self-reported drug use by youths**

**CITATION:** Gfroerer, J. (1985). Influence of privacy on self-reported drug use by youths. In B. A. Rouse, N. J. Kozel, & L. G. Richards (Eds.), *Self-report methods of estimating drug use: Meeting current challenges to validity* (HHS Publication ADM 85-1402, NIDA Research Monograph 57, pp. 22-30). Rockville, MD: National Institute on Drug Abuse.

**PURPOSE/OVERVIEW:** Underreporting of drug use by survey respondents has always been a major concern of drug abuse survey researchers. This concern particularly applies when the respondents are youths because they might fear being punished if their use of drugs was discovered by parents. By 1985, a number of studies had been conducted to determine whether underreporting was a serious problem and to identify procedures that could be used to obtain the most valid data. Although some of the results were contradictory, most studies concluded that reliable, valid self-reported drug use data could be obtained (Hubbard et al., 1976; O'Malley et al., 1983; Single et al., 1975; Smart, 1975; Smart & Jarvis, 1981). Factors identified as possibly affecting the reporting of drug use by youths included the type of questionnaire (interview vs. self-administered), characteristics of the interviewer, the degree of anonymity of the respondent, the setting (home vs. school), and the degree of privacy during the interview (Johnston, in press; Sudman & Bradburn, 1974). In several studies, self-administered questionnaires were shown to produce higher reported prevalence of drug use than interviews. Although many factors that may affect underreporting could be controlled by researchers, it was not possible to achieve complete privacy (i.e., no one else in the room) in every interview when conducting a household survey. Given this limitation, it was important to assess the impact of the lack of privacy on the results of a survey, both to assess the potential impact on the validity of data from that survey and also to provide general information on the importance of privacy for future surveys.

**METHODS:** This study used data from a national probability sample of youths 12 to 17 years of age to assess the impact of the lack of privacy on the results of surveys. Depending on household composition, interviews were conducted with one adult only, one youth only, or both an adult and a youth. The surveys collected data on whether respondents had used various licit and illicit drugs in the past month, past year, or ever. Measures were taken to ensure privacy and confidentiality.

**RESULTS/CONCLUSIONS:** It was found that interviewers were successful in obtaining privacy, and the degree of privacy obtained was the same in a 1979 interview as in a 1982 interview. In general, it was found that most population groups that reported higher prevalence of drug use did not have significantly more privacy than lower prevalence groups. One exception was whites, who had slightly higher reported prevalence of cigarette and alcohol use than other races and also had more privacy during interviews. An investigation of this indicated that the

privacy effect is independent of race. In conclusion, it appeared that reporting of drug use by youths was affected by the degree of privacy during the interview, even when a self-administered answer sheet was used by the respondent. This underscored the importance of achieving maximum privacy when conducting drug surveys and raised questions regarding the validity of data from surveys in which adequate privacy was not obtained.

### **The validity of self-reported drug use data: The accuracy of responses on confidential self-administered answer sheets**

CITATION: Harrell, A. V., Kapsak, K. A., Cisin, I. H., & Wirtz, P. W. (1986, December). *The validity of self-reported drug use data: The accuracy of responses on confidential self-administered answer sheets* (Contract No. 271-85-8305). Rockville, MD: National Institute on Drug Abuse.

PURPOSE/OVERVIEW: The purpose of this study was to examine the validity of self-reported data on the use of illicit drugs, including marijuana, cocaine, hallucinogens, and heroin, as well as the nonmedical use of some psychotherapeutic drugs. The focus of the study was the criterion validity of self-reported data and the extent to which respondent accounts of their drug use could be shown to be consistent with an independent indicator (i.e., clinic records of respondents' drug use).

METHODS: To examine the validity of self-reported drug use data, a sample of clients discharged from drug treatment facilities was interviewed using a questionnaire and interviewing procedures developed for the 1985 National Household Survey on Drug Abuse (NHSDA). Respondents' self-reported drug use was compared with the criterion of clinic records on the drugs being abused by these same respondents at the time of their admission to treatment. To isolate relatively recent drug use that would be less subject to failures of recall, the sample included only respondents who had been admitted to treatment within a year of the interview. To avoid possible bias in the drug use reports that could occur if respondents were aware that their answers could be verified, respondents were treated exactly as though they had been randomly selected to participate in NHSDA, and no reference was made to any advance knowledge of their drug use history. Similarly, interviewers were unaware of the fact that respondents had been treated for drug abuse, both to protect the privacy of the respondents and to prevent interviewer expectations from influencing responses to the questionnaire.

The interview provided self-reported data on the use of a wide variety of illicit drugs and on respondents' symptoms of, and treatment for, drug abuse. The clinic records provided data on the drugs (up to three) that were considered problems at the time of admission to treatment. The validity "tests" were whether respondents reported on the survey use of the drugs listed in the clinic records at the time of admission in the past year or at any time in the past. Another validity "test" was whether the respondents reported receiving any treatment for the use of drugs other than alcohol during the past year or at any time in the past. The analysis investigated the relationship of reporting accuracy to demographic characteristics and to factors that might influence the respondents' willingness or ability to respond accurately, such as the level of privacy during the interview and the amount of time between admission to the program and the interview.

**RESULTS/CONCLUSIONS:** In general, the reporting pattern on the drug use questions was consistent with the thesis that the higher the level of perceived discrimination (i.e., social stigma) associated with a behavior, the more prevalent denial of the behavior will be. Variation in the level of privacy during the interview did not appear to influence the willingness of respondents to report drug use. Responses to questions about drug treatment were generally less accurate than the drug use self-reported data.

### **The nominative technique: A new method of estimating heroin prevalence**

**CITATION:** Miller, J. D. (1985). The nominative technique: A new method of estimating heroin prevalence. In B. A. Rouse, N. J. Kozel, & L. G. Richards (Eds.), *Self-report methods of estimating drug use: Meeting current challenges to validity* (HHS Publication No. ADM 85-1402, NIDA Research Monograph 57, pp. 104-124). Rockville, MD: National Institute on Drug Abuse.

**PURPOSE/OVERVIEW:** In 1985, the nominative technique was a relatively new method of indirect survey-based estimation that was being developed expressly for the purpose of estimating heroin prevalence in the general population. This technique, which involves asking respondents to report on their close friend's heroin use, is essentially an attempt to reap the benefits of survey research, while at the same time avoiding some of the major problems of the self-report method. The primary purpose of the nominative technique is to minimize respondent denial of socially undesirable behavior. Another possible advantage is achieving coverage of "hard-to-reach" deviant population groups. The nominative question series was inserted in the 1977, 1979, and 1982 National Household Surveys on Drug Abuse (NHSDAs) (Miller et al., 1983).

**METHODS:** The resulting nominative estimates of heroin prevalence are presented and contrasted with corresponding self-report estimates.

**RESULTS/CONCLUSIONS:** Nominative estimates of heroin prevalence have been consistently higher than self-reports of heroin use. During this time, nominative data have generally followed mainstream patterns of drug use: Nominative estimates for young adults and for males were higher than nominative estimates for older individuals, youths, and females. Moreover, in 1985, the recent downward trends in drug use were replicated by the nominative heroin data. Thus, the overall picture presented by the nominative data—similar patterns but higher levels of prevalence—seemed to support the validity of the new approach. Nevertheless, considerable caution should be exercised in interpreting nominative data. This is chiefly because a substantial minority of nominators cannot report the number of other close friends of the heroin user who also "know." Although missing data have been handled by a conservative imputation rule, the fact that so many individuals are unable to provide an answer to this key question casts doubt on the accuracy of the answers that were given. In fact, the nominative approach might tend to produce overestimates because of the potential for undercounts of the numbers of others who "know."

## 1990–1991

### **Generalized standard error models for proportions in complex design surveys**

CITATION: Bieler, G. S., & Williams, R. L. (1990). Generalized standard error models for proportions in complex design surveys. In *Proceedings of the 1990 American Statistical Association, Survey Research Methods Section, Anaheim, CA* (pp. 272-277). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: Generalized variance functions often are employed when large numbers of estimates are to be published from a survey. Generalized variance functions lessen the volume of published reports where presentation of each standard error estimate would essentially double the size of the tabular presentations. In addition, generalized functions may facilitate secondary data analyses that were not conducted in the initial publications. Generalized variance functions also may provide more stable estimates of variance by diminishing the variability of the individual variance estimates themselves.

METHODS: In this paper, the authors studied some generalized models for the standard error of proportion estimates from the 1988 National Household Survey on Drug Abuse. A log-linear model based on the concept of a design effect was presented. The final model was evaluated against the simple average design effect (DEFF) model.

RESULTS/CONCLUSIONS: The authors concluded that, for these data, the domain-specific average DEFF model and the simple log-linear model both provided adequate generalized standard error models. They were surprised that the log-linear model including only effects for  $\log(p)$ ,  $\log(1-p)$ , and  $\log(n)$  performed so well. They expected that domain effects would be required in the model to account for the differential average cluster sizes by domain for this multistage sample design. It appeared that the slope for  $\log(n)$  in the model being different from the simple random sampling (SRS) value of  $-0.5$  accounted for most of the cluster size effect.

### **The feasibility of collecting drug abuse data by telephone**

CITATION: Gfroerer, J. C., & Hughes, A. L. (1991). The feasibility of collecting drug abuse data by telephone. *Public Health Reports*, 106(4), 384-393.

PURPOSE/OVERVIEW: This article examines the feasibility of using telephone surveys to collect data on illicit drug use. It compares estimates on drug use from three data collection efforts, all carried out in 1988. The first, an in-person survey, was the 1988 National Household Survey on Drug Abuse (NHSDA). The other two surveys were telephone surveys, the Quick Response Survey (QRS) and the 1988 Texas Survey on Substance Use Among Adults.

METHODS: To assess the potential effects of noncoverage of nontelephone households on survey estimates, data from the 1988 NHSDA were used to compare estimates for those in telephone households with those in nontelephone households on lifetime and past year use of marijuana and cocaine. Estimates on illicit drug use and demographic items also were compared

between the 1988 NHSDA and the two telephone surveys. Comparable survey items were asked on all three surveys. For comparisons with the telephone surveys, NHSDA responses were edited to be consistent with questionnaire skip patterns in these surveys. For comparisons with estimates from the Texas Survey on Substance Use Among Adults, cases were restricted to the metropolitan areas of three certainty primary sampling units (PSUs) from the NHSDA. NHSDA data then were reweighted to match age group by metropolitan area totals from the Texas survey.

**RESULTS/CONCLUSIONS:** Within the NHSDA itself, drug use prevalence rates among those in telephone households were much lower than those in nontelephone households although most differences were not statistically significant. Estimates on drug use from QRS were consistently lower than those from NHSDA, even after controlling for differences in sample composition (QRS respondents reported higher levels of income and education than NHSDA respondents). For the combined metropolitan areas in Texas, one statistically significant difference was found between estimates from the NHSDA and the Texas survey on lifetime use of marijuana.

### **Improving the comprehension of reference periods**

**CITATION:** Hubbard, M. L., Lessler, J. T., Graetz, K. A., & Forsyth, B. H. (1990). Improving the comprehension of reference periods. In *Proceedings of the 1990 American Statistical Association, Survey Research Methods Section, Anaheim, CA* (pp. 474-479). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** This paper describes a pair of studies directed at examining alternative questioning strategies for the National Household Survey on Drug Abuse (NHSDA). The NHSDA provides comprehensive national data on drug use. As of 1990, its data were collected using a combination of interviewer- and self-administered questions. General questions on health and demographics were interviewer administered, while questions on more sensitive topics, such as the use of illegal drugs, employed self-administered answer sheets. Respondents frequently demonstrated a variety of errors when answering these types of questions. For example, in the 1988 survey, about 20 percent of the sample indicated they had used a drug at least once in the previous 12 months yet, in a subsequent question, indicated they had not used the drug in the previous 12 months.

**METHODS:** Under contract with the National Institute on Drug Abuse (NIDA), RTI designed a number of methodological studies exploring possible cognitive explanations and remedies for errors of this sort.

**RESULTS/CONCLUSIONS:** The results of these experiments were somewhat disappointing. Many surveys ask respondents to recall the number of events in specific reference periods. Because both bounded and intensive interviewing is expensive and time-consuming, it was hoped that these studies would demonstrate the effectiveness of less intensive techniques. This was not the case, although, in Study 2, the results were in the predicted direction. Perhaps respondents' estimates of the frequency of personal drug use were not generated by directly recalling individual incidences but by an estimation or averaging process that was not sensitive to the time periods used in these studies. If this was the case, it was possible that the ability to recall other types of events would be affected by using these types of less intensive procedures to

anchor the reference periods. However, it also may have been the case that either a more intensive interaction than was used in these studies was needed to anchor the reference periods or that respondents who were not in the anchoring condition were independently generating these kinds of anchors.

## **Analysis of survey data on drug experience by mode of data collection**

**CITATION:** Hughes, A. L., & Gfroerer, J. C. (1990). Analysis of survey data on drug experience by mode of data collection. In *Proceedings of the 1990 American Statistical Association, Survey Research Methods Section, Anaheim, CA* (pp. 401-406). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** In the late 1980s, the increasing concern over the drug abuse problem in the United States created a need for more data on the nature and extent of drug abuse. Policymakers demanded timely, accurate data at the national, State, and local levels to guide them in directing programs and funding toward the goal of reducing drug abuse and to measure progress in these programs. At the national level, the primary source of data on the prevalence of illicit drug use had been and remained the National Household Survey on Drug Abuse (NHSDA), sponsored by the National Institute on Drug Abuse (NIDA). The NHSDA was a probability sample survey of the U.S. household population that employed personal visit interviews with all selected respondents. The high cost of conducting household surveys raised interest in using telephone survey methodology to collect drug use prevalence data. In fact, several States had conducted drug use surveys by telephone and compared their State results with national NHSDA data (Frank, 1985; Spence, Fredlund, & Kavinsky, 1989). Some studies comparing data from telephone surveys to personal visit surveys showed that comparable health-related data (Massey, Barker, & Moss, 1979) and sociodemographic data (Monsees & Massey, 1979) could be obtained from the two methods. Other studies showed that sensitive data from telephone surveys were inferior to data collected from personal interviews (see Section 2 of this paper). However, very little research had been done investigating how the quality of data collected from drug use surveys differed by the two methods. The purpose of this paper was to present an analysis of the Quick Response Survey (QRS) and NHSDA in an attempt to evaluate the feasibility of using telephone survey methodology to collect data on illicit drug use.

**METHODS:** In 1988, NIDA funded a telephone survey on drug abuse through a cooperative agreement with the Food and Drug Administration (FDA) using a QRS contract. The QRS was conducted by Chilton Research Services at about the same time the 1988 NHSDA was in the field.

**RESULTS/CONCLUSIONS:** National estimates of illicit drug use from a telephone survey (QRS) and a personal visit survey (NHSDA) were compared. Overall, the results showed that estimates of illegal drug use from the QRS were significantly biased downward—a bias large enough to seriously distort confidence interval estimates. There were considerable differences in sample and drug use characteristics by phone status in the NHSDA data alone; however, because about 93 percent of the 1988 household population had phones, estimates from NHSDA telephone households were similar to estimates from the total sample. The effect of leading into sensitive questions on the use of illicit drugs by first asking about legal drugs cannot be ignored.

QRS respondents may have been more willing to admit their illegal drug use if they had been eased into these sensitive questions with less sensitive questions about such substances as cigarettes, alcohol, and prescription-type drugs, as is done on the NHSDA. This also could have had the negative effect of reduced response rates due to the length of the interview; however, although the authors concluded that this contextual issue needed further study, it was unlikely that this difference could have accounted for the substantial differences in reported drug use in the two surveys.

# 1992

## Follow-up of nonrespondents in 1990

CITATION: Caspar, R. A. (1992). Follow-up of nonrespondents in 1990. In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (HHS Publication No. ADM 92-1929, pp. 155-173). Rockville, MD: National Institute on Drug Abuse.

PURPOSE/OVERVIEW: This chapter presents the results of a special follow-up survey of individuals who did not respond to the 1990 National Household Survey on Drug Abuse (NHSDA). The aim was to understand the reasons people chose not to participate, or were otherwise missed in the survey, and to use this information in assessing the extent of the bias, if any, that nonresponse introduced into the 1990 NHSDA estimates.

METHODS: To assess the impact of nonresponse, a follow-up study was undertaken on a subset of nonrespondents to the 1990 survey. The study was conducted in the Washington, DC, area, a region with a traditionally high nonresponse rate. The follow-up survey design included a \$10 incentive and a shortened version of the instrument. The response rate for the follow-up survey was 38 percent.

RESULTS/CONCLUSIONS: The results of the follow-up study did not demonstrate definitively either the presence or absence of a serious nonresponse bias in the 1990 NHSDA. In terms of demographic characteristics, follow-up respondents appeared to be similar to the original NHSDA respondents. Estimates of drug use for follow-up respondents showed patterns that were similar to the regular NHSDA respondents. Only one statistically significant difference was found between the two groups for a composite measure of drug use (including cigarettes and alcohol) at any time during their lives, with follow-up respondents reporting higher rates than regular NHSDA respondents. Follow-up respondents also reported higher rates of lifetime use of cocaine, marijuana, alcohol, and cigarettes than regular NHSDA respondents, although the differences were not statistically significant. On the other hand, follow-up respondents had lower reported rates of past month use of cigarettes, alcohol, marijuana, and any drug use than regular NHSDA respondents, but these differences were not statistically significant. Another finding was that among those who participated in the follow-up survey, one third were judged by interviewers to have participated in the follow-up because they were unavailable for the main survey request. Finally, 27 percent were judged to have been swayed by the incentive and another 13 percent were judged to have participated in the follow-up due to the shorter instrument.

## **A follow-up study of nonrespondents to the 1990 National Household Survey on Drug Abuse**

CITATION: Caspar, R. A. (1992). A follow-up study of nonrespondents to the 1990 National Household Survey on Drug Abuse. In *Proceedings of the 1992 American Statistical Association, Survey Research Section, Boston, MA* (pp. 476-481). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: As of 1992, little was known about the drug use patterns of individuals who were nonrespondents to the National Household Survey on Drug Abuse (NHSDA). Age, gender, race, and Hispanic origin were known from screening information for individual nonrespondents, but these data provided only minimal guidance in assessing the drug use patterns of individuals who were not directly included in the regular survey estimates. To the extent that nonrespondents differed from respondents in their drug use and to the extent that NHSDA nonresponse adjustment procedures failed to take account of this difference, estimates from the NHSDA were subject to nonresponse bias. The issue of potential nonresponse bias was not a trivial one. The overall interview nonresponse rate in the 1990 NHSDA was 18 percent, with considerably higher rates in many locales. In the Washington, DC, metropolitan area, for example, the nonresponse rate was 27 percent.

METHODS: To assess the impact of such nonresponse, a follow-up study was undertaken of a subset of nonrespondents to the 1990 survey. For logistical reasons, the study was conducted in a single metropolitan area with a relatively high nonresponse rate. By offering nonrespondents a shortened questionnaire and monetary incentive, the authors hoped to persuade as many as possible to participate in the follow-up study. Their aim was to understand the reasons people chose not to participate—or were unavailable to participate in the survey—and to use this information in assessing the extent of the bias, if any, that nonresponse introduced into the 1990 NHSDA estimates.

RESULTS/CONCLUSIONS: The results of the follow-up study did not definitively demonstrate either the presence or the absence of a serious nonresponse bias in the 1990 NHSDA. For reasons of cost, the follow-up study was confined to the Washington, DC, metropolitan area, and the results may not have been generalizable to other areas of the country. Similarly, with a response rate of 38 percent, there remained a sizable majority of sample nonrespondents for whom no information was obtained. Anecdotal information from follow-up interviewers suggested that these hard-core nonrespondents may have differed significantly in their drug use behaviors from individuals interviewed as part of either the regular NHSDA or the follow-up study. In terms of demographic characteristics, follow-up respondents appeared to be similar to the original NHSDA respondents. Estimates of drug use for follow-up respondents showed patterns similar to the regular NHSDA respondents. Only one statistically significant difference was found between the two groups—for the composite measure of drug use at any time during their lives.

From both the qualitative and the quantitative data presented here, it would appear that the NHSDA nonrespondents who were interviewed in the follow-up study were quite similar to the respondents interviewed as part of the regular NHSDA data collection. Interviewers working on the follow-up noted that individuals who continued to refuse to participate appeared to have

something to hide and to be afraid of answering questions about drugs. Whether this was indicative of higher rates of drug use among hard-core nonrespondents was unknown. Adding the follow-up cases to the regular NHSDA sample made little difference to the NHSDA estimates of drug use prevalence. The authors did not know, however, how convincing these remaining hard-core nonrespondents would affect NHSDA's estimates. Because the authors could not follow up all nonrespondents, the assessment of nonresponse bias provided by this method was, of necessity, incomplete. Nevertheless, it could indicate the potential impact on NHSDA prevalence estimates of alternative survey procedures (e.g., selective use of monetary incentives) to increase response rates.

### **Inconsistent reporting of drug use in 1988**

CITATION: Cox, B. G., Witt, M. B., Traccarella, M. A., & Perez-Michael, A. M. (1992). Inconsistent reporting of drug use in 1988. In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (HHS Publication No. ADM 92-1929, pp. 109-153). Rockville, MD: National Institute on Drug Abuse.

PURPOSE/OVERVIEW: This chapter describes the inconsistencies that occurred in the reporting of drug use in the 1988 National Household Survey on Drug Abuse (NHSDA). A methodological study was conducted to locate and measure the faulty data in the 1988 NHSDA; results are summarized in this chapter. The study examined inconsistent responses within and among various sections of the questionnaire used for the 1988 NHSDA.

METHODS: For this analysis, the authors used data taken from a partially edited data file. Many of the inconsistent responses included in the analyses reported in this chapter could be eliminated by logical editing of the data. This methodological analysis was designed to identify those aspects of the survey questionnaire and survey process that caused difficulty for participants. It was determined that these aspects of the survey, in turn, should become primary targets for revision in future administrations of the NHSDA.

RESULTS/CONCLUSIONS: The first section of this chapter provides an overall indication of approximately how many respondents made errors, where the errors were made, and the demographic distribution of the respondents who made the errors. Subsequent sections present further details in the inconsistencies in reporting found on the alcohol, marijuana, and cocaine forms, and in reporting concomitant drug use. Findings include the following: (1) A total of 24 percent of those who reported using at least one drug in their lifetime answered at least one question on drug use inconsistently. Among respondents who reported using at least one drug in the past 12 months, 31 percent answered at least one question on drug use inconsistently. (2) Inconsistencies occurred most frequently on the alcohol form, although on a percentage basis, inconsistencies in alcohol reporting occurred at lower rates than for many less frequently used substances. (3) No inconsistencies occurred within the cigarette form, which, unlike other forms, was administered by the interviewer. The same was true for the sedative, tranquilizer, stimulant, and analgesic forms, which provided an option to skip the form if they had never used these drugs. (4) Young respondents, those between the ages of 12 to 17, were less likely to provide inconsistent responses than older respondents.

## **A method for identifying cognitive properties of survey items**

**CITATION:** Forsyth, B. H., & Hubbard, M. L. (1992). A method for identifying cognitive properties of survey items. In *Proceedings of the 1992 American Statistical Association, Survey Research Methods Section, Boston, MA* (pp. 470-475). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** Question wording can lead to error and bias survey measurement (e.g., Forsyth & Lessler, 1991a; Groves, 1989; Sudman & Bradburn, 1974; Turner & Marlin, 1984). This paper reports on research to develop and test a method for identifying survey items that are difficult for respondents to answer due to their cognitive demands. For example, items may be difficult to answer if the wording is difficult to understand, if response requires detailed memory recall, or if response categories fail to cover the range of respondent experience. The authors' aim was to design a taxonomy of item characteristics that could be used to identify potentially problematic survey items. The research reported was part of the methodological study of the National Household Survey on Drug Abuse (NHSDA) sponsored by the National Institute on Drug Abuse (NIDA). Major goals of the larger research project were to (1) identify potential sources of measurement error in NHSDA, (2) revise survey materials and survey procedures that seem to contribute to avoidable measurement error, and (3) test revisions and identify improved strategies for measuring patterns drug use. This paper focuses on reducing measurement errors that arise from the cognitive processes that respondents use when answering survey items.

**METHODS:** The appraisal results summarized here were used as one basis for identifying method improvements to test under more formal laboratory and field test conditions. Based in part on these appraisal results, the authors developed three sets of improvements. First, they used laboratory and field test procedures to investigate test decomposition approaches for defining technical terminology and complex reference sets. Second, they used laboratory and field test methods to test procedures for anchoring reference periods. Third, they used field test methods to test experimental questionnaire materials that eliminated hidden questions by using branching instructions and skip patterns.

**RESULTS/CONCLUSIONS:** As reported in other papers in the 1992 volume, the experimental and field test results suggested that the authors' appraisal methodology made an important contribution to identifying sources of response inconsistencies, response biases, and response variability. Additional research is necessary before this coding scheme can be used as a general purpose tool for analyzing survey items. The authors have worked to clarify, refine, and trim their coding categories, collapsing some while expanding upon others. In addition, they indicate a need for further research to provide valid tests of the coding scheme once it has been refined. Although further development and testing is necessary, the authors believe that they have begun to develop a cost-effective method for systematizing expert evaluations and for identifying and cataloging critical aspects of questionnaire wording and format.

## **An overview of the National Household Survey on Drug Abuse and related methodological research**

CITATION: Gfroerer, J. (1992). An overview of the National Household Survey on Drug Abuse and related methodological research. In *Proceedings of the 1992 American Statistical Association, Survey Research Methods Section, Boston, MA* (pp. 464-469). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: This paper provides a brief description of the history of the National Household Survey on Drug Abuse (NHSDA) and a summary of methodological research carried out prior to 1992.

METHODS: A summary of the methodological research that had been done in conjunction with NHSDA is given.

RESULTS/CONCLUSIONS: The results summarized in this paper include the following: (1) A field test conducted in 1990 found lower rates of drug use reporting for an interviewer-administered survey than self-administered surveys, particularly for more recent use. The magnitudes of the differences varied by substance reported (smallest for alcohol, larger for marijuana and even larger for cocaine). Differences in estimates between the interviewer- and self-administered forms were larger for youths than adults. (2) A skip pattern experiment conducted in 1992 found that prevalence rates for marijuana and cocaine were lower in a version that employed skip patterns than the main NHSDA sample (which did not employ skip patterns). (3) A follow-up survey of nonrespondents in the Washington, DC, area for the 1990 NHSDA had a 38 percent response rate. The most frequent reason offered for participating in the follow-up survey was that the respondent was unavailable for the initial survey request. (4) A validation study in which a sample of clients recently discharged from drug treatment facilities were administered the NHSDA interview found higher levels of underreporting on the survey (relative to information from treatment records) for more socially deviant behaviors (e.g., higher levels of underreporting for heroin and cocaine, lower levels of underreporting for marijuana). The study also found very low levels of reporting drug treatment on the interview (only 38 percent). (5) Finally, additional research summarized here found that reporting of drug use by youths is more likely under conditions of complete privacy.

## **The incidence of illicit drug use in the United States, 1962–1989**

CITATION: Gfroerer, J., & Brodsky, M. (1992). The incidence of illicit drug use in the United States, 1962-1989. *British Journal of Addiction*, 87(9), 1345-1351.

PURPOSE/OVERVIEW: As of the early 1990s, epidemiological drug studies over the prior four decades had primarily focused on the prevalence of illicit drug use and not on the incidence of use as well as trends. In this paper, retrospective drug use data from the National Household Survey on Drug Abuse (NHSDA) were used to provide estimates for the number of new drug users each year from 1962 to 1989.

**METHODS:** Data from 5 years of the NHSDA were combined to yield a total sample size of 58,705 respondents. The variables of interest in the surveys were date of birth, date of interview, and reported age at first use for each drug (heroin, cocaine, marijuana, and hallucinogens), which allowed the researchers to pinpoint an exact date of first drug use for each respondent. The methods and validity of this approach also were evaluated.

**RESULTS/CONCLUSIONS:** The findings show that the peak year for new marijuana users was in 1973 and remained roughly stable until about 1980, at which point it started to drop. The average age of new users remained around the late teens since 1962. Patterns for hallucinogen use were similar to marijuana use. Peak estimates for incidence of cocaine use occurred in 1982 and subsequently had seen a dropoff. Retrospective estimates for drug use incidence have some limitations in that they exclude drug users who died before 1985, which could be significant for the estimates prior to 1962. Retrospective reports also are affected by recall bias in that estimates for the early years relied on a recall of over 20 years. However, advantages of the retrospective estimates were that a large sample size was able to provide stability in the estimates.

### **Introduction. In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies***

**CITATION:** Gfroerer, J. C., Gustin, J., & Turner, C. F. (1992). Introduction. In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (HHS Publication No. ADM 92-1929, pp. 3-10). Rockville, MD: National Institute on Drug Abuse.

**PURPOSE/OVERVIEW:** This is the introduction to a monograph that presents a range of studies assessing the accuracy of alternative methods for survey measurement of drug use. This volume reports the results of a program of methodological research designed to evaluate and improve the accuracy of measurements made in the National Household Survey on Drug Abuse (NHSDA), the Nation's major survey for monitoring drug use. Most of the research reported in this volume began in 1989 and was conducted by a team of scientists from the Research Triangle Institute and the National Institute on Drug Abuse.

**METHODS:** This introductory chapter describes the research components and briefly outlines the origins and purposes of the NHSDA and its methodological research program. The final section of this chapter describes the organization of the volume and the contents of the chapters.

**RESULTS/CONCLUSIONS:** N/A.

### **Collecting data on illicit drug use by phone**

**CITATION:** Gfroerer, J. C., & Hughes, A. L. (1992). Collecting data on illicit drug use by phone. In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (HHS Publication No. ADM 92-1929, pp. 277-295). Rockville, MD: National Institute on Drug Abuse.

**PURPOSE/OVERVIEW:** This chapter evaluates the feasibility of using telephone surveys to collect data on illicit drug use. It compares estimates on drug use from three data collection efforts, all carried out in 1988: (1) the 1988 National Household Survey on Drug Abuse (NHSDA), (2) the Quick Response Survey (QRS), and (3) the 1988 Texas Survey on Substance Use Among Adults.

**METHODS:** Using data from the 1988 NHSDA, estimates for those in telephone households were compared with those in nontelephone households on lifetime and past year use of marijuana and cocaine in order to examine the effects of noncoverage of nontelephone households in a telephone survey. Estimates on illicit drug use and demographic items from the 1988 NHSDA also were compared with those from QRS, after NHSDA data were reedited for consistency with QRS skip patterns. Finally, estimates from the 1988 NHSDA from Texas were compared with those from the Texas Survey on Substance Use Among Adults. For this analysis, data were restricted to the metropolitan areas of three certainty primary sampling units (PSUs) from NHSDA. NHSDA estimates were combined with those from the QRS and then these in turn were compared with estimates from the Texas survey.

**RESULTS/CONCLUSIONS:** Although some differences were not statistically significant, drug use prevalence among telephone households was generally much lower than that among nontelephone households in NHSDA. Estimates on drug use from QRS were consistently lower than those from NHSDA, even after accounting for differences in sample composition (QRS respondents reported higher levels of income and education than NHSDA respondents). One statistically significant difference was found (among four items examined) between estimates from NHSDA for the combined metropolitan areas in Texas and the Texas survey (for lifetime use of marijuana).

### **The intersection of drug use and criminal behavior: Results from the National Household Survey on Drug Abuse**

**CITATION:** Harrison, L., & Gfroerer, J. (1992). The intersection of drug use and criminal behavior: Results from the National Household Survey on Drug Abuse. *Crime and Delinquency*, 38(4), 422-443.

**PURPOSE/OVERVIEW:** The purpose of this paper is to explore the relationship between drug use and criminal behavior focusing on the difference between licit and illicit drug use.

**METHODS:** In 1991, questions about criminal involvement and criminal behavior were added to the National Household Survey on Drug Abuse (NHSDA). A total of 32,594 respondents completed the NHSDA interview, which contained self-report questions on drug use and criminality. The relationship between drug use and crime was analyzed using cross-tabulations and logistic regression.

**RESULTS/CONCLUSIONS:** The results showed a positive correlation between criminal behavior and higher levels of drug use. Respondents who did not report using drugs or alcohol during the past year had the lowest levels of criminal behavior. After controlling for sociodemographic characteristics, such as age, gender, and race, higher drug use still was

associated significantly with criminal behavior. The results showed further findings by the type of criminal behavior and other individual-level covariates.

### **Effects of decomposition of complex concepts**

CITATION: Hubbard, M. L., Pantula, J., & Lessler, J. T. (1992). Effects of decomposition of complex concepts. In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (HHS Publication No. ADM 92-1929, pp. 245-264). Rockville, MD: National Institute on Drug Abuse.

PURPOSE/OVERVIEW: This chapter reports the results of the authors' experiment and compares the responses obtained using an alternative measurement strategy with those obtained with the National Household Survey on Drug Abuse (NHSDA) in the early 1990s.

METHODS: In the experiment, the authors tested a questionnaire in which complex concepts were decomposed into a number of simpler elements for which more straightforward questions could be formulated.

RESULTS/CONCLUSIONS: The evidence with regard to measurements of the nonmedical use of psychotherapeutics indicated that decomposing the NHSDA question on this issue substantially increased the reporting of such use. The results were striking, particularly for painkillers. The estimated prevalence of nonmedical use of painkillers more than doubled when the authors decomposed the concept into constituent parts and asked respondents separate questions about each part. For stimulants and tranquilizers, the differences were not as striking as for painkillers, but they were substantial. The analysis of responses to the identical question using this new questioning strategy suggested that some respondents may have used a personal definition of nonmedical use regardless of the instructions provided on the questionnaire. It became clear to the authors that survey designers cannot rely on respondents to follow instructions to classify usage as medical or nonmedical.

### **1992 National Household Survey on Drug Abuse: Findings of first quarter skip test: Final report**

CITATION: Lessler, J. T., & Durante, R. C. (1992, October 7). *1992 National Household Survey on Drug Abuse: Findings of first quarter skip test: Final report* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 271-91-5402, RTI 5071). Research Triangle Park, NC: Research Triangle Institute.

PURPOSE/OVERVIEW: In 1990, the National Institute on Drug Abuse (NIDA) conducted a large methodological field test of the National Household Survey on Drug Abuse (NHSDA), primarily to evaluate the effect of using interviewer-administered versus self-administered questionnaires. In this study, some questionnaire answer sheets incorporated skip patterns. It was found that, in general, respondents were able to follow skip patterns that were not too complex, particularly if skips were always to the top of a page. Generally, when errors occurred, they resulted in respondents answering additional questions unnecessarily, so there was no loss of

data. However, the design of the field test did not allow a determination of the effect of skip patterns on reporting of drug use.

In an attempt to address this, NIDA conducted an experiment during the first 3 months of 1992 to test a new questionnaire that incorporated skip patterns into the drug use answer sheets. An experimental questionnaire was developed that included a number of skip patterns that allowed respondents to skip out of entire sections of questions if they responded "no" to an initial question on whether they had used a drug. This questionnaire was called the skip version. The regular NHSDA questionnaire was called the nonskip version. Differences between the two questionnaires varied by section in the questionnaire.

**METHODS:** The Skip Pattern Experiment was embedded in the first quarter 1992 NHSDA sample. One eighth of the first quarter sample was randomly assigned to receive the skip questionnaire, while the other seven eighths received the nonskip version. Assignment of questionnaire versions to sampled dwelling units was done in advance of any contact by field staff. Allocation of the skip version was done within sample segments to maximize the power of statistical comparisons between the two groups. Interviewers were trained to administer both versions of the questionnaire. Overall, the nonskip version was administered to 7,149 respondents, and the skip version was administered to 974 respondents.

**RESULTS/CONCLUSIONS:** Overall, this methodological study indicated that using skip patterns tended to reduce the prevalence of illicit drug use. There also was an indication that the bias due to using skips would not be uniform across different populations, as it seemed to be more pronounced among respondents with more education. It was not possible to conclude whether the lack of privacy or the desire to avoid additional questions was operating. However, it is interesting to note that the skip in the alcohol questions had no apparent impact. This would suggest that the sensitivity of the illicit drug use questions and the loss of privacy in the skip version are most important.

### **Effects of mode of administration and wording on reporting of drug use**

**CITATION:** Turner, C. F., Lessler, J. T., & Devore, J. W. (1992). Effects of mode of administration and wording on reporting of drug use. In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (HHS Publication No. ADM 92-1929, pp. 177-219). Rockville, MD: National Institute on Drug Abuse.

**PURPOSE/OVERVIEW:** This chapter assesses the impact of question wording and mode of administration on estimates of prevalence of drug use from the National Household Survey on Drug Abuse (NHSDA).

**METHODS:** The authors describe the design of the methodological field test conducted in 1990 and review estimates of the prevalence of self-reported drug use obtained by the different questionnaire versions used in this field test.

**RESULTS/CONCLUSIONS:** Although the results were not always consistent for every substance examined, on balance, the results indicate that having interviewers administer the

questionnaire reduces the reporting of drug use. This conclusion is supported by the finding that lack of privacy during an interview had a negative effect on the reporting of drug use, particularly for respondents 12 to 17 years of age for whom a parent is likely to be present. The authors' finding that self-administered forms yield more reports of drug use does not appear to be due to a greater number of marking errors.

### **Effects of mode of administration and wording on data quality**

**CITATION:** Turner, C. F., Lessler, J. T., George, B. J., Hubbard, M. L., & Witt, M. B. (1992). Effects of mode of administration and wording on data quality. In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (HHS Publication No. ADM 92-1929, pp. 221-243). Rockville, MD: National Institute on Drug Abuse.

**PURPOSE/OVERVIEW:** In this 1992 chapter, the authors review the impact of various factors on the completeness and consistency of the reporting of drug use on the National Household Survey on Drug Abuse (NHSDA).

**METHODS:** The authors summarize findings of their investigations of selected aspects of the quality of the data produced by the different versions of the survey questionnaire and the different modes of administration used in a 1990 NHSDA field test. The authors consider the rates of nonresponse, the extent to which respondents and interviewers correctly executed the branching instructions embedded in the questionnaires, and the internal consistency of the reports of drug use given by respondents.

**RESULTS/CONCLUSIONS:** The results of the authors' analyses provide the key ingredients for conclusions concerning the relative quality of the data provided by the different versions of the survey evaluated in the 1990 NHSDA field test. Respondents were, in general, capable of responding to a self-administered form, even when that form included many branching or skip instructions. The changes made to the NHSDA questionnaire generally improved respondent understanding of the questions and thereby improved the quality of the data collected.

### **Survey measurement of drug use: Methodological studies**

**CITATION:** Turner, C. F., Lessler, J. T., & Gfroerer, J. C. (Eds.). (1992). *Survey measurement of drug use: Methodological studies* (HHS Publication No. ADM 92-1929). Rockville, MD: National Institute on Drug Abuse.

**PURPOSE/OVERVIEW:** This 1992 volume presents a range of studies assessing the accuracy of alternative methods for the survey measurement of drug use. It contains demonstrations of the liability of such measurements in response to variation in the measurement procedures that are employed. The editors urge that these demonstrations should not be taken as an indictment of the use of surveys to measure drug use. The data obtained from a measurement process are properly seen to be a joint function of the phenomenon under study and the protocol used to make the measurement. This joint parentage is reflected in the frequent practice of reporting such

measurements by reference to both the measurement outcome and the protocol used in the measurement.

**METHODS:** The volume is divided into six parts, the first of which is the introduction. Part II contains two chapters that present the results of cognitive research on the National Household Survey on Drug Abuse (NHSDA) questionnaire. Part III comprises three chapters that analyze different aspects of past NHSDAs to identify problems. Part IV reports the results of a large-scale field experiment that tested the new version of the NHSDA questionnaire. Part V presents the results of two studies that complement the field experiment. Part VI is the single concluding chapter that summarizes major findings.

**RESULTS/CONCLUSIONS:** The various chapters in this volume demonstrate the appropriateness of such practices in the reporting and use of survey measurements of drug use. The authors offer a wide-ranging view of the impact of measurement procedures on survey measurements of drug use, and they introduce new techniques for diagnosing problems with survey questionnaires and for designing improved ones.

### **Item nonresponse in 1988**

**CITATION:** Witt, M. B., Pantula, J., Folsom, R. E., & Cox, B. G. (1992). Item nonresponse in 1988. In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (HHS Publication No. ADM 92-1929, pp. 85-108). Rockville, MD: National Institute on Drug Abuse.

**PURPOSE/OVERVIEW:** This chapter examines the patterns of missing data in the 1988 National Household Survey on Drug Abuse (NHSDA). The quality of a survey instrument may be gauged in part by the completeness and internal consistency of the data it produces. Survey instruments that produce more and more consistent measurements are preferable. Evidence of incomplete or inconsistent responses to a survey can be used to identify the aspects of a survey that are in need of remediation. Respondents' failure to answer questions may, for example, reflect their misunderstanding of the survey instructions or bafflement with a particular question.

**METHODS:** The authors assessed the levels and patterns of nonresponse to questions in the 1988 NHSDA. The nonresponse was not strongly concentrated on any one form, nor focused on any one type of question. Nevertheless, a few general statements were made about the nonresponse. A disproportionately large number of respondents were people who never used the drug that was discussed on the form. A great deal of nonresponse resulted from those questions that asked about an individual's concomitant drug use and those questions that asked the age at which a respondent first took a particular drug.

**RESULTS/CONCLUSIONS:** The authors concluded that the high rate of nonresponse to questions regarding age occurred because people had a hard time recalling what had happened in the past. The substantial nonresponse on concomitant drug use may be due to the question's sensitivity.

*This page intentionally left blank*

# 1993

## **Self-reported drug use data: What do they reveal?**

CITATION: Harrison, E. R., Haaga, J., & Richards, T. (1993). Self-reported drug use data: What do they reveal? *American Journal of Drug and Alcohol Abuse*, 19(4), 423-441.

PURPOSE/OVERVIEW: The purpose of this article was to measure the accuracy of self-reported drug use by examining the consistency of the self-reports over time.

METHODS: Using data from the 1985 and 1990 National Household Surveys on Drug Abuse (NHSDAs), trends in self-reported drug use for marijuana and cocaine were compared across time. Two years of data from the National Longitudinal Survey Youth Cohort (NLS-Y) also were analyzed to observe any discrepancies in reported drug use over time.

RESULTS/CONCLUSIONS: The results from the NHSDA analysis revealed a declining trend in drug use starting in the late 1980s, which was consistent with data collected from other self-report studies. A cohort analysis from the NHSDA data revealed that fewer people in each birth cohort reported ever using marijuana or cocaine in 1990 than in 1985. This phenomenon occurred particularly in the birth cohorts born in the mid-1950s to the mid-1960s. The decrease in reported lifetime drug use over time led researchers to conclude that either people were not as willing to report drug use as they used to be or else people's perceptions of what constitutes drug use had changed. Results from the NLS-Y data were consistent with the NSDUH results in that 20 percent of the respondents in the 1984 survey who admitted to ever using drugs denied using them in the 1988 survey. The most significant indicators of discrepancy in reported drug use were being African American and the frequency of drug use. These results raised the question of whether outside factors, such as public disapproval or drug use, affect respondents' willingness to reveal their prior drug use.

## **Report on 1990 NHSDA-Census Match**

CITATION: Parsley, T. (1993). *Report on 1990 NHSDA-Census Match* (RTI 4469-15, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 271-89-8333). Research Triangle Park, NC: Research Triangle Institute.

PURPOSE/OVERVIEW: In 1991, the National Institute on Drug Abuse (NIDA) and the Research Triangle Institute (RTI) cooperated with the U.S. Bureau of the Census on a study of nonresponse in surveys. This effort was part of the multiagency work conducted by Robert M. Groves and Mick P. Couper at the Census Bureau to study nonresponse in seven major Federal Government surveys, including the National Household Survey on Drug Abuse (NHSDA). The study involved linking data from a sample of respondents and nonrespondents to the 1990 NHSDA with their 1990 decennial census data to provide descriptive information about NHSDA nonrespondents. NHSDA and census records were linked by Census Bureau clerks using primarily address and other location information.

**METHODS:** A total of 5,030 NHSDA households were selected to be matched to 1990 census records. All 860 screener nonresponse households were selected. In addition, all 1,821 households with at least one interview nonrespondent were selected for matching. To allow comparisons between respondents and nonrespondents on the census items, a random systematic sample of 1,938 households was selected in which all sample individuals completed the interview. Finally, to assess the accuracy of interviewers' classifications, a sample of 411 cases classified as vacant, temporary dwellings, or nonhousing units by the NHSDA interviewer was selected.

**RESULTS/CONCLUSIONS:** The Census Match Study demonstrates that response rates are not constant across various interviewer, respondent, household, and neighborhood characteristics. To the extent that rates of drug use vary by these same characteristics, bias due to nonresponse may be a problem. However, it is not always the case that low response rates occur in conjunction with high drug use prevalence. Some populations with low response rates (e.g., older adults and high-income populations) tend to have low rates of drug use. On the other hand, some populations (e.g., residents of large metropolitan areas and men) have low response rates and high drug use rates. In estimating overall prevalence, many of these potential sources of bias would be in opposite directions and would therefore tend to cancel each other.

### **Evaluation of results from the 1992 National Household Survey on Drug Abuse (NHSDA)**

**CITATION:** Peer Review Committee on National Household Survey on Drug Abuse (NHSDA) [James T. Massey, Chair; Marc Brodsky; Tom Harford; Lana Harrison; Dale Hitchcock; Ron Manderscheid; Nancy Pearce; Marilyn Henderson; Beatrice Rouse; & Ron Wilson]. (1993, June 3). *Evaluation of results from the 1992 National Household Survey on Drug Abuse (NHSDA)* [memo analyzing the decrease in drug prevalence among the black population between the 1992 NHSDA and previous NHSDAs]. Research Triangle Park, NC: Research Triangle Institute.

**PURPOSE/OVERVIEW:** The preliminary results from the first two quarters of 1992 raised some concerns at the Research Triangle Institute (RTI) and the Substance Abuse and Mental Health Services Administration (SAMHSA) because of some differences between the 1992 results and the survey results from 1988, 1990, and 1991. The main concern was a greater than expected decline in reported drug use in 1992, primarily among blacks. Particularly puzzling was a drop in the reported lifetime use of illicit drugs, which would not be expected to change from one year to the next. Because of the survey differences between years, the Office of Applied Studies (OAS) formed a Peer Review Committee (PRC) to evaluate the results from the 1992 National Household Survey on Drug Abuse (NHSDA) and to make recommendations about the release and publication of the results.

**METHODS:** The first step in the evaluation process was to review the key results from the 1988 to 1992 NHSDAs and to identify the key differences and any inconsistencies in trends from 1988 through 1992. The committee reached the same conclusions that had been reached by the OAS and RTI staffs. Namely, the 1992-reported drug use for blacks was consistently lower than the 1991 results. A number of possible causes for the 1992 decline in drug use were explored by

analyzing the results from the 1988 to 1992 surveys, although some of the possible explanations could be evaluated only indirectly or subjectively.

**RESULTS/CONCLUSIONS:** The consensus of the committee was that the observed differences between 1991 and 1992 cannot be explained by a single factor, although several small differences were found among the factors examined. In conducting its analysis of the NHSDA, the committee concluded that the design and procedures for sampling, weighting, editing, and imputing the survey results were statistically sound. Great care had been taken by the National Institute on Drug Abuse (NIDA), OAS, and RTI to implement the survey procedures and to evaluate the quality of the results. The committee concluded that the unexpected decrease in lifetime drug use among blacks was an example of what can occasionally occur in survey estimates, particularly when a large number of different estimates are generated and comparisons are made. Often, a review of the procedures will uncover an error in the process. In other cases, such as the NHSDA, an explanation for the unexpected results may never be found.

## **Drug use measurement: Strengths, limitations, and recommendations for improvement**

**CITATION:** U.S. General Accounting Office. (1993). *Drug use measurement: Strengths, limitations, and recommendations for improvement* (GAO/PEMD-93-18, Report to the Chairman, Committee on Government Operations, House of Representatives, Program Evaluation and Methodology Division). Washington, DC: U.S. Government Printing Office.

**PURPOSE/OVERVIEW:** Policymakers, researchers, and planners must have accurate drug use information if they are to properly assess the Nation's current drug prevalence pattern and trends. However, the quality of data can be constrained by methodological problems, available research technology, and environmental and budgetary limitations. Because millions of dollars are spent on drug prevalence studies, it is important to evaluate the current state of drug use measurement practices within a cost feasibility context. In response to a request by the Chairman of the House Committee on Government Operations, the GAO investigated the issue of drug use measurement by (1) reporting on three nationally prominent drug studies, (2) assessing the methodological strengths and limitations of each of these studies, and (3) developing recommendations for the improvement of drug prevalence estimates.

**METHODS:** The GAO examined the National Household Survey on Drug Abuse (NHSDA), the High School Senior Survey (HSSS), and the Drug Use Forecasting (DUF) study of booked arrestees. The GAO evaluated the methodological strengths and limitations of these three studies in terms of the degree to which their research operations satisfied generally accepted criteria. The GAO also developed guidelines for improving drug prevalence estimates, focusing particularly on high-risk groups.

**RESULTS/CONCLUSIONS:** NHSDA is a sophisticated study of drug use patterns and trends, but is limited by the exclusion of groups at high risk for drug use, problematic measurement of heroin and cocaine use, and reliance on subject self-reports. HSSS has similar limitations. Therefore, both surveys provide conservative estimates of drug use. DUF cannot be generalized to booked arrestees in the geographic areas sampled. The GAO found that drug prevalence

estimates could be improved and money could be saved if NHSDA and HSSS were administered in alternate years. The GAO cited several ways of validating these two surveys and estimating the extent of underreporting. Expanding the subsamples of current surveys and conducting new studies aimed at hard-to-reach, high-risk groups should improve the coverage of underrepresented target populations.

# 1994

## **Repeated measures of estimation of measurement bias for self-reported drug use with applications to the National Household Survey on Drug Abuse**

CITATION: Biemer, P., & Witt, M. (1994). Repeated measures of estimation of measurement bias for self-reported drug use with applications to the National Household Survey on Drug Abuse. In *Proceedings of the National Institute on Drug Abuse technical review: The validity of self-reported drug use*. Rockville, MD: National Institute on Drug Abuse.

PURPOSE/OVERVIEW: Direct estimates of response bias in self-reports of drug use in surveys require that essentially error-free determinations of drug use be obtained for a subsample of survey respondents. The difficulty of obtaining determinations that are accurate enough for estimating validity is well documented in the literature. Methods such as specimen (hair, urine, etc.) analysis, proxy reports, and the use of highly private and anonymous modes of interview all have to contend with error rates that may be only marginally lower than those of the parent survey. Thus, any methodology for direct validity estimation must rely to some extent on approximations and questionable assumptions.

METHODS: In this article, the authors considered a number of methods that rely solely on repeated measures data to assess response bias. Because the assumptions associated with these approaches do not require highly accurate second determinations, they may be more easily satisfied in practice. One such method for bias estimation for dichotomous variables that is considered in some detail provides estimates of misclassification probabilities in the initial measurement without requiring that the second measure be accurate or even better than the first. This methodology does require, however, that two subpopulations exist that have different rates of prevalence but whose probabilities of false-positive and false-negative error are the same. The applicability of these methods for self-reported drug use is described and illustrated using data from the National Household Survey on Drug Abuse (NHSDA). In the discussion of the results, the importance of these methods for assessing the validity of self-reported drug use is examined.

RESULTS/CONCLUSIONS: In this paper, a general model for studying misclassification in self-reported drug use was presented, and the model then was extended to the case where two measurements of the same characteristic are available for a sample of respondents. For the two measurements case, the general model required seven parameters while only three degrees of freedom were available for estimation. It is shown how the assumptions typically made for test-retest, true value, improved value, and Hui and Walter methods relate to the general model.

## Estimating substance abuse treatment need from the NHSDA

CITATION: Epstein, J. F., & Gfroerer, J. C. (1994). Estimating substance abuse treatment need from the NHSDA. In *Proceedings of the 1994 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Toronto, Ontario, Canada* (pp. 559-564). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: By the early 1990s, several methods had been developed to estimate treatment need using the National Household Survey on Drug Abuse (NHSDA). These methods were generally developed to estimate drug abuse treatment need, not alcohol treatment need. The National Institute on Drug Abuse (NIDA) developed an illicit drug index in 1989 that defined heavy drug users who need treatment as those who had used illicit drugs at least 200 times in the past year. An estimated 4 million Americans met this criteria in 1988. Another model developed by NIDA was based on reported problems and symptoms of abuse or dependence on illicit drugs. This method was an approximation of criteria in the 3rd edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) published by the American Psychiatric Association (APA, 1980). This method produced an estimate of 3.5 million people with drug abuse or dependence in 1991. A study sponsored by the Institute of Medicine (IOM) in 1990 developed a methodology for estimating drug abuse treatment need using a combination of frequency of use and problems/symptoms indicators from the NHSDA. The IOM also supplemented these NHSDA estimates with data from prison and homeless populations to produce a more comprehensive estimate of need. This method resulted in an estimated 4.6 million people needing treatment in 1988. Throughout this paper, this method is referred to as the FREQ method. As of 1994, a method for estimating treatment need from the NHSDA had been considered by the Office of Applied Studies (OAS) of the Substance Abuse and Mental Health Services Administration (SAMHSA). This method used an algorithm that approximated the criteria from the 3rd revised edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R) for illicit drug dependence (APA, 1987). This method is referred to as the DEP method.

METHODS: This paper compared the FREQ method and the DEP method. Because the FREQ method did not apply to alcohol, alcohol was not included in the comparison. One objective of this comparison was to better understand the strengths and weaknesses of these two methods, providing a basis for further improvements in treatment need estimation by SAMHSA. This comparison is based on an analysis of the 1991 NHSDA.

RESULTS/CONCLUSIONS: From the investigation of the DEP and FREQ estimates of treatment need on different populations, the following conclusions were reached: (1) Many populations of serious drug users were only partially assigned as needing treatment by both the FREQ and the DEP methods. For example, less than half of the individuals who used needles in the past year, less than half of those who used heroin in the past year, and less than half of those who used cocaine weekly or more often were estimated to need treatment by each method. (2) Both methods estimated a very small proportion of individuals using prescription drugs as needing treatment. (3) The two methods did not estimate the same people as needing treatment (kappa statistics were less than .5 for all populations and less than .2 for more than half the populations). (4) The FREQ method generally estimated a larger proportion of individuals as needing treatment. In summary, the FREQ method defined a broader group of drug abusers as

needing treatment, while the DEP method defined a smaller more severely impaired group of drug abusers as needing treatment. For some populations clearly in need of treatment, the FREQ estimate was better than the DEP estimate because it estimated a larger proportion of the individuals in these populations as needing treatment. For other populations, the DEP estimate may differentiate better between the less serious and more serious drug users clearly in need of treatment.

## **Small area estimation for the National Household Survey of [sic] Drug Abuse**

CITATION: Folsom, R. E., & Liu, J. (1994). Small area estimation for the National Household Survey of [sic] Drug Abuse. In *Proceedings of the 1994 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Toronto, Ontario, Canada* (pp. 565-570). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: In this paper, the authors summarized their planned approach for producing small area estimates of drug and alcohol use for selected U.S. States and metropolitan statistical areas (MSAs). The small area statistics of primary interest were population prevalence estimates of illicit drug and alcohol dependency as ascertained from responses to National Household Survey on Drug Abuse (NHSDA) questionnaire items. The authors produced separate rates for 32 demographic subpopulations defined by the three-way cross-classification of gender with four age intervals and four race/ethnicity groups where the three non-Hispanic categories excluded Hispanics. Although the authors expected most of the small area statistics for the "other races" category to be suppressed due to excessive mean squared error (MSE) estimates, the inclusion of this other races category permitted them to report statistics for Hispanics, whites, and blacks employing the same race/ethnicity definitions typically used in government data sources.

METHODS: The authors' small area estimation (SAE) strategy began by predicting block group-level drug or alcohol dependency/use rates for all the 1990 block groups in a State or MSA small area by demographic domain (the 32 gender by age by race/ethnicity subpopulations). The authors used a logistic regression predictor for the drug or alcohol dependency/use rate of domain- $d$  in block group- $k$  for MSA/county- $j$  in State- $i$ .

RESULTS/CONCLUSIONS: To quantify the uncertainty associated with the authors' nested random effect logistic model estimators, they linearized the logit-transformed versions of the statistics. For States where the fraction of the domain- $d$  population that resides in sample counties was negligible, the linearized form of the equation was made proportional to a formula. To approximate another formula, the authors employed the posterior covariance matrices specified above with estimated  $D$  replacing  $D$ . To account for the significant additional variation that may result, the authors derived a special case of Prashad and Rao's (1991) result for their three-level nested random effects model.

## **Prevalence of drug use in the DC metropolitan area household and nonhousehold populations: 1991**

CITATION: National Institute on Drug Abuse. (1994). *Prevalence of drug use in the DC metropolitan area household and nonhousehold populations: 1991* (Technical Report #8, Washington, DC, Metropolitan Area Drug Study, DC\*MADS). Rockville, MD: National Institute on Drug Abuse, Division of Epidemiology and Prevention Research. [Prepared by Bray, R. M., Kroutil, L. A., Wheelless, S. C., Marsden, M. E., & Packer, L. E.]

PURPOSE/OVERVIEW: This study from the early 1990s examined the prevalence of illicit drug, alcohol, and tobacco use among members of household and nonhousehold populations and a combined aggregate population aged 12 or older in the District of Columbia metropolitan statistical area (DC MSA). In addition, selected characteristics of three drug-abusing subgroups in the household and aggregate populations were examined: crack cocaine users, heroin users, and needle users. The study had three methodological objectives: (1) to investigate the effect that combining data from household and nonhousehold populations had on estimates of the prevalence of drug use and number of users, (2) to determine whether the addition of nonhousehold populations allowed more detailed demographic analyses to be conducted for specific drug-using behaviors, and (3) to identify important methodological issues when combining and analyzing data from household and nonhousehold populations.

METHODS: Household population data were collected as part of the DC MSA oversample of the 1991 National Household Survey on Drug Abuse (NHSDA) (Inter-university Consortium for Political and Social Research [ICPSR] 6128). Nonhousehold population data were subsetted from the 1991 DC\*MADS Institutionalized Population Study and the 1991 DC\*MADS: Homeless and Transient Population Study (ICPSR 2346).

RESULTS/CONCLUSIONS: Household survey topics included age at first use, as well as lifetime, annual, and past month usage for the following drug classes: marijuana and hashish, cocaine (and crack), hallucinogens, heroin, inhalants, alcohol, tobacco, anabolic steroids, nonmedical use of prescription drugs (including psychotherapeutics), and polysubstance use. Respondents also were asked about substance abuse treatment history, problems resulting from use of drugs, perceptions of the risks involved, personal and family income sources and amounts, need for treatment for drug or alcohol use, mental health and access to care, illegal activities and arrests, and needle-sharing. Demographic data for the household population included gender, race, age, ethnicity, marital status, motor vehicle use, educational level, job status, income level, veteran status, and past and current household composition. Topics from the Institutionalized Study and the Homeless and Transient Population Study included history of homelessness, alcohol and drug treatment or counseling, illegal activities and arrests, physical health, pregnancy, mental health, mental health care, employment, income and expenditures, living arrangements and population movement, and specific and general drug use. Drugs covered included tobacco, alcohol, marijuana and hashish, inhalants, cocaine and crack, hallucinogens, heroin, stimulants, and tranquilizers. Data also provided information on age at first use, route of administration, polysubstance use, perceived risks, and insurance coverage. Demographic data for the nonhousehold population included gender, age, marital status, race, education, military service, and number of children and other dependents.

## **Breach of privacy in surveys on adolescent drug use: A methodological inquiry**

CITATION: Schütz, C. G., Chilcoat, H. D., & Anthony, J. C. (1994). Breach of privacy in surveys on adolescent drug use: A methodological inquiry. *International Journal of Methods in Psychiatric Research*, 4(3), 183-188.

PURPOSE/OVERVIEW: Drug use is one of the more sensitive topics of surveys in the United States. Although there is much effort to ensure the confidentiality of surveys, there are some occasions where this confidentiality may be breached. For example, when interviews are done in the home, a third party, such as a parent, may be present throughout the interview. The authors investigated whether this breach of privacy during National Household Survey on Drug Abuse (NHSDA) interviews affected the reporting of drug use among adolescents who participated in the survey.

METHODS: The authors used NHSDA data to test the association between the level of privacy and the reported prevalence of drug use. They selected survey participants aged 12 to 17 years old ( $n = 2,152$ ) for the analysis. Interviewers provided ratings of privacy after the interview.

RESULTS/CONCLUSIONS: It was found that adolescent respondents were less likely to report their involvement with tobacco, alcohol, and illegal substances, such as marijuana and cocaine, when their parents were present during the interview than when complete privacy was reported. However, when there was a nonparent present during the interview, adolescent respondents were more likely to report their use of alcohol, tobacco, or illicit drugs than when there was complete privacy. However, in additional analyses that controlled for other variables, it was found that the presence of someone other than a parent did not have statistically significant effects on drug use reporting.

## **Ratio estimation of hard-core drug use**

CITATION: Wright, D., Gfroerer, J., & Epstein, J. (1994). Ratio estimation of hard-core drug use. In *Proceedings of the 1994 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Toronto, Ontario, Canada* (pp. 541-546). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: In the early 1990s, the need for accurate estimates of the size of the so-called "hard-core" drug-using population was substantial. Regardless of how it was specifically defined, this population of heavy drug users was likely to need significant resources for treatment of their drug problem and associated medical and other problems. Hard-core drug users also had been shown to be responsible for a disproportionate amount of crime.

This paper describes a method for estimating the prevalence of hard-core drug use based on the National Household Survey on Drug Abuse (NHSDA) in conjunction with outside sources and the methodology of ratio estimation. In ratio estimation, one can often obtain a better estimate of a population total if there is a known population total for a related variable. Then the estimate of

the total is  $X' = (x/y)*Y$ , where  $x$  is the variable of interest,  $y$  is the related variable, and  $Y$  is the known population total for the related variable.

Another way of describing this method is to say that it "inflates" (i.e., gives more weight to) the drug prevalence data from the NHSDA for populations with characteristics that are known to be related to hard-core drug use but also are underestimated. In this case, it is known that the NHSDA undercounted arrestees and drug treatment populations, so the authors "ratio adjust" the NHSDA hard-core drug use estimates upward to externally derive counts of arrestees and treatment clients that are believed to be accurate.

In survey sampling theory, ratio estimation often is associated with the desire to improve the precision of an estimate. The ratio estimate will be better, in the sense that it will have a smaller variance when certain conditions are met. (See Section 6 of this 1994 paper on precision of the estimates.) However, in this application, the authors were less interested in variance reduction and more interested in bias reduction. Ratio estimates have been used for a number of years to adjust for nonresponse and to adjust to known population counts, often based on a census. This application represents an extension of those earlier uses to one in which a known population count was used to adjust NHSDA sample estimates for underreporting and undercoverage.

**METHODS:** The information that the authors made use of is the count of the number of individuals in treatment centers for drug abuse during the past year (1992) from the National Drug and Alcoholism Treatment Unit Survey (NDATUS) (Substance Abuse and Mental Health Services Administration [SAMHSA], 1993b) and the known count of the number of arrests (for any crime other than minor traffic violations) during the past year (1991) from the Federal Bureau of Investigation's Uniform Crime Reports (UCRs) (Maguire, Pastore, & Flanagan, 1993).

**RESULTS/CONCLUSIONS:** A complete evaluation and comparison of the ratio estimation procedure with other methods of estimating hard-core drug use was beyond the scope of this paper. However, the authors offered some overall statements about ratio estimation: (1) Ratio estimation does not fully account for underreporting and undercoverage in the NHSDA. In particular, for the population not arrested and not in treatment, the method does not adjust for underreporting at all. Thus, the authors considered these estimates of hard-core drug use to be improvements on the generally published NHSDA estimates (using the simple expansion estimator) but still conservative estimates. (2) The ratio estimation model, as applied in this case, relied primarily on regularly updated and consistently collected data from the NHSDA, NDATUS, and UCR, and a relatively small number of easily understood assumptions. Thus, it was likely to be able to provide more reliable trend information (given constant levels of underreporting) than the previously used methods, which relied more heavily on assumptions that could change over time. (3) Because the model relied primarily on the NHSDA sample design and weighting, it was possible to develop estimates of the variances of ratio-adjusted estimates. This was generally not possible in the methods previously used.

The authors indicated that there were three primary areas for further investigation:

(1) the population counts, (2) the assumptions made about the ratios used, and (3) a search for "unbiased" methods to estimate the ratio.

# 1995

## **Estimation of drug use incidence using NHSDA**

**CITATION:** Johnson, R., Gerstein, D., & Gfroerer, J. (1995). Estimation of drug use incidence using NHSDA. In *Proceedings of the 1995 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Orlando, FL* (pp. 437-442). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** This paper reviews the methods of data collection and statistical estimation used to estimate rates of drug use incidence in the National Household Survey on Drug Abuse (NHSDA), the primary source of data on the prevalence of drug use in the United States, and assesses possible biases in these estimated rates. NHSDA is conducted annually by the Substance Abuse and Mental Health Services Administration (SAMHSA) and is a continuing, cross-sectional, personal interview survey of individuals aged 12 or older in the civilian, noninstitutionalized population of the United States.

**METHODS:** This paper uses data from the 1988, 1990, 1991, 1992, and 1993 NHSDAs. There were approximately 8,000 respondents in each of the 1988 and 1990 NHSDAs and approximately 30,000 in each of the 1991, 1992, and 1993 NHSDAs. The five surveys used essentially the same drug use questions, and the 1991 to 1993 surveys oversampled the same six major metropolitan areas.

**RESULTS/CONCLUSIONS:** Rates of drug use incidence are critical statistics in monitoring trends in drug use in the United States. A review of data collection and estimation procedures in the NHSDA suggested three sources of possible bias in such rates: (1) differential mortality, (2) memory errors, and (3) social acceptability/fear of disclosure. The young ages of most first drug users and historical data on U.S. mortality suggested differential mortality was unlikely to be a serious source of bias in rates computed for reference periods fewer than 30 years ago. Except for the cigarette questions, NHSDA used self-administered answer sheets for all drug questions and other procedures designed to reduce social acceptability bias and respondents' perceived risks of disclosure. (The 1994 NHSDA converted the tobacco section to the self-administered format.) Preliminary analyses comparing estimates for the same reference periods computed using NHSDAs conducted in different years suggested that biases due to memory errors were small for reference periods prior to the 1990s.

*This page intentionally left blank*

# 1996

## **Repeated measures estimation of measurement bias for self-reports of drug use with applications to the National Household Survey on Drug Abuse**

CITATION: Biemer, P., & Witt, M. (1996). Repeated measures estimation of measurement bias for self-reports of drug use with applications to the National Household Survey on Drug Abuse. *Journal of Official Statistics*, 12(3), 275-300.

PURPOSE/OVERVIEW: This article explores a variety of methods used to assess the validity of self-reported drug use. Using repeated measures, such as re-interview methods, test-retest, record check studies, and biological test validation, the researchers estimated measurement bias on self-reported drug use. This paper describes an improved method for estimating false-positive and false-negative self-reports in repeated measures studies.

METHODS: The authors explain that traditional repeated measures methods rely on the assumption that the second measure has considerably lower error rates than the parent survey; however, that is often not the case. In this paper, a number of methods that do not require that the second measurement be more accurate than the first were evaluated. One such methodology, the Hui-Walter method, is implemented on the National Household Survey on Drug Abuse (NHSDA) where repeated measures on several variables are available. The goal was to estimate the percentages of false negatives and false positives on several key variables to assess the validity of self-reported drug use.

RESULTS/CONCLUSIONS: This article presents a general model for estimating the proportion of misclassification in self-reported drug studies. The model was applied to studies where repeated measures of the same item are available. The Hui-Walter method produced estimates for the prevalence of drug use that are adjusted for the rate of false positives and false negatives. The next step is to use the Hui-Walter method in validating self-report errors in studies that use a biological test as a validation measure.

## **Aggregating survey data on drug use across household, institutionalized, and homeless populations**

CITATION: Bray, R. M., Wheelless, S. C., & Kroutil, L. A. (1996). Aggregating survey data on drug use across household, institutionalized, and homeless populations. In R. Warnecke (Ed.), *Sixth Conference on Health Survey Research Methods: Conference proceedings* [Breckenridge, CO] (HHS Publication No. PHS 96-1013, pp. 105-110). Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

PURPOSE/OVERVIEW: Since 1971, the National Household Survey on Drug Abuse (NHSDA) series has provided key information about the extent of drug use among people living in households in the United States. However, some NHSDA estimates have been criticized in

recent years because the survey has been limited in its ability to adequately represent populations that are at high risk for abusing alcohol or illicit drugs, such as incarcerated or homeless people. This paper describes an effort to combine data obtained from members of households, institutionalized populations, and homeless populations aged 12 or older into an aggregate population in the District of Columbia metropolitan statistical area (DC MSA). The paper also presents findings about drug use prevalence for both household and aggregate populations.

**METHODS:** This research is based on data from three surveys in the DC MSA: the DC MSA oversample of the 1991 NHSDA, the DC\*MADS Institutionalized Study, and the DC\*MADS Homeless and Transient Population Study. Because the populations included in these studies are generally defined in terms of residence, the overlap is very small: less than 0.5 percent of the total combined population. However, adjustments were made to avoid multiple counting of the subpopulations when producing aggregate estimates. Respondents were first classified according to the number of overlapping surveys for which they could have been selected. At most, the overlap occurred only in two of the three surveys. Analysis weights were adjusted for these respondents. The adjusted weights were summed with the final analysis weights to form multiplicity-adjusted weights for the aggregate data.

**RESULTS/CONCLUSIONS:** To assess the trade-offs in the use of multiplicity estimates, the variance of key estimates was examined. The authors found that the aggregate dataset produced unbiased estimates of the prevalence of illicit drug, alcohol, and cigarette use among the eligible population in the DC MSA. The authors also found that combining data from household and nonhousehold populations resulted in slightly higher prevalence estimates than from household data alone, although it had little impact on estimates within demographic subgroups for any illicit drug, alcohol, and cigarette use. However, the aggregate population totals yielded more illicit drug users. Findings from this study may underscore a potential limitation in reporting overall estimates for a large population. Such general estimates can obscure high rates of drug use or related problems among subgroups that constitute only a small percentage of the overall population.

### **Increasing response rates and data quality in personal interview surveys without increasing costs: An application of CQI to the NHSDA**

**CITATION:** Burke, B., & Virag, T. (1996). Increasing response rates and data quality in personal interview surveys without increasing costs: An application of CQI to the NHSDA. In *Proceedings of the 1996 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Chicago, IL, Vol. 2* (pp. 1026-1031). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** The consensus among survey researchers in 1996 was that there had been a decline in recent years in response rates for large, national personal visit surveys. Many national surveys required increased data collection efforts to maintain acceptable response rates (i.e., rates comparable with previous rounds [Groves & Couper, 1992]). As part of Research Triangle Institute's (RTI) continuing effort to change this trend, RTI developed a program of continuous quality improvement (CQI) for its field staff. This paper presents the history, design, implementation, and results of the first year of this CQI program that was implemented for the

1995 National Household Survey on Drug Abuse (NHSDA). Many of the features of its CQI program are readily transferable to other national personal visit surveys.

**METHODS:** In the summer of 1994, RTI convened a task force of individuals drawn from RTI management, NHSDA project management, survey methodologists, statisticians, and programmers/analysts to investigate data collection improvement areas for the NHSDA. The task force held a series of meetings and several focus group sessions involving the NHSDA field supervisors. Based on these discussions, RTI developed a program of CQI for NHSDA field interviewing. The objectives of this program were fourfold: (1) to increase screening and interviewing response rates to their highest levels ever; (2) to reduce costs and increase data quality; (3) to provide the field supervisors (FSs) and the field interviewers (FIs) with the tools, resources, and support to achieve unprecedented gains in quality; and (4) to appropriately reward those FSs and FIs who were responsible for the improvements in proportion to their contributions to the program's objectives.

To achieve these objectives, NHSDA management staff planned a system of performance measurement, communication, and improvement that emphasized teamwork, numerical goals, and data-driven decision making. An important feature of this approach was a program of recognition and reward commensurate with team performance on the key, critical success factors that define field interviewing quality. This was implemented through a team-building concept called "Together Everyone Achieves More" (TEAM), which incorporated basic Total Quality Management and CQI concepts.

**RESULTS/CONCLUSIONS:** The authors saw some encouraging results at the end of the first year of the TEAM program. They had seen the best response rates since 1992 as well as a reduction in field costs per completed interview from the levels experienced during the 1994 NHSDA. They determined that continued evaluation of performance on the remainder of the 1995 NHSDA and throughout the 1996 NHSDA would allow them to evaluate whether the TEAM program would continue to show increases in response rates along with decreases in operating costs. The authors concluded that more research was required to determine whether this approach to managing national personal interview surveys could be applied to other studies outside the NHSDA project.

## **Social environmental impacts on survey cooperation**

**CITATION:** Couper, M. P., & Groves, R. M. (1996). Social environmental impacts on survey cooperation. *Quality and Quantity*, 30(2), 173-188.

**PURPOSE/OVERVIEW:** This paper explores the effects on survey cooperation rates due to the social context and environment of the respondent household, such as urbanicity, population density, crime, and social disorganization.

**METHODS:** Using nonresponse data from six surveys, including the National Household Survey on Drug Abuse (NHSDA), and comparing them with census data, social context, and environmental factors were analyzed among 11,600 cases. Multivariate modeling was performed using logistic regression to analyze the difference between responders and nonresponders.

**RESULTS/CONCLUSIONS:** The results confirm the authors' hypotheses that highly urban areas, such as densely populated metropolitan areas, and high crime rates were associated with higher refusal rates. Controlling for household-level variables, such as household structure, race, age, number of children, and socioeconomic status, strengthened the effect of social environmental factors, but it did not eliminate them. A large percentage of the influence of environmental factors on response rates can be attributed to household-level variables. These findings show that survey participation is associated with both person- and household-level factors.

### **Special populations, sensitive issues, and the use of computer-assisted interviewing in surveys**

**CITATION:** Gfroerer, J. (1996, April). Special populations, sensitive issues, and the use of computer-assisted interviewing in surveys. In R. Warnecke (Ed.), *Sixth Conference on Health Survey Research Methods: Conference proceedings* [Breckenridge, CO] (HHS Publication No. PHS 96-1013, pp. 177-180). Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

**PURPOSE/OVERVIEW:** This discussion paper summarizes a session at the Sixth Conference on Health Survey Research Methods, which included papers on three separate but related areas of interest to health survey researchers: surveys of special populations, collecting data on sensitive issues, and the use of computer-assisted interviewing (CAI). Data from the National Household Survey on Drug Abuse (NHSDA) were used in the research to address the collection of sensitive data from youths. Each of these three topics received increasing attention in the early to mid-1990s.

Surveys of special populations have become more important as health planners and policymakers required data to address the health care needs of specific population subgroups. Although sometimes these data could be obtained from ongoing broad-based surveys, often it is necessary to conduct separate surveys targeting special populations. Available ongoing surveys may not have sufficient numbers of cases in the population of interest or may not even include the population in their universe (e.g., most household surveys excluded homeless people). Many of the same methodological issues apply whether the special population is surveyed in a limited study or as part of a larger survey with broader coverage.

Surveys of sensitive issues have become more prevalent, in part to provide critical data describing emerging health problems, such as acquired immunodeficiency syndrome (AIDS) and drug abuse. These health problems require survey researchers to collect data on sensitive topics, such as sexual behavior and illegal activities. This in turn requires new and innovative methods for ensuring the validity of the data collected in surveys.

The third major area discussed in this session is the use of CAI, which was rapidly becoming the standard for all large-scale surveys by the mid-1990s. Many studies document the data collection, processing, and quality benefits of CAI. As costs continue to decline and improved technology and software become available, many surveys have been converting to CAI,

including computer-assisted telephone interviewing (CATI), computer-assisted personal interviewing (CAPI), computer-assisted self-interviewing (CASI), and audio computer-assisted self-interviewing (ACASI).

**METHODS:** The emergence of each of these three major areas in health survey research occurred somewhat independently and for unrelated reasons. However, it is difficult to discuss them separately. The methodological questions that they generate often overlap, and the answers given by research in these areas are sometimes meaningful only in the context of the others. For example, some topics that are sensitive for some special populations may not be sensitive for others. An example of this is alcohol use, which is thought to be sensitive for underage youths but not for adults. Similarly, some types of CAI (e.g., CASI) may work very well for most populations but may be problematic for some special populations, requiring specialized methods. And respondents' willingness to report sensitive data may vary with different types of CAI.

**RESULTS/CONCLUSIONS:** During the period of rapid conversion of surveys to CAI in the 1990s, it was critical that methodological research include studies of the benefits and effects on data that CAI has in surveys of special populations and on sensitive topics. The six papers presented in this session add important new knowledge to this growing body of research, and are useful to government agencies conducting surveys of special populations and on sensitive issues and that are considering the use of CAI.

### **The influence of parental presence on the reporting of sensitive behaviors by youth**

**CITATION:** Horm, J., Cynamon, M., & Thornberry, O. (1996, April). The influence of parental presence on the reporting of sensitive behaviors by youth. In R. Warnecke (Ed.), *Sixth Conference on Health Survey Research Methods: Conference proceedings* [Breckenridge, CO] (HHS Publication No. PHS 96-1013, pp. 141-145). Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

**PURPOSE/OVERVIEW:** A continuing challenge facing survey researchers in the mid-1990s was administering questions on respondent-perceived sensitive behaviors in a manner that minimized response bias. Negative personal behaviors could be frequently underreported or inaccurately reported because of fear of disclosure, including inadvertent disclosure to household members. Telephone and self-administered interviews, which were often used as solutions to this problem, had serious shortcomings (Aquilino, 1994; Gfroerer, 1985; Groves, 1990; Johnson, Hougland, & Clayton, 1989; Schwarz, Strack, Hippler, & Bishop, 1991). Telephone coverage was incomplete, and respondents may not have been comfortable answering questions about sensitive behaviors when there was a concern that another household member may overhear or listen in on the conversation. Further, the absence of a social relationship between the interviewer and the respondent during telephone interviews was believed to reduce the respondent's willingness to reveal personal information. Although generally producing higher reported levels of sensitive behaviors, a self-administered document had the major shortfall of requiring the respondent to have a sophistication and skill in both reading and filling out complex questionnaires. In the mid-1990s, the recently developed computer-assisted

self-interviewing (CASI) technique, with audio, overcame some of the problems of literacy and privacy, but it was not always feasible due to budgetary constraints (Mosher, Pratt, & Duffer, 1994; O'Reilly, Hubbard, Lessler, Biemer, & Turner, 1994). This paper discusses an alternative mode that was inexpensive and offered privacy for household surveys on sensitive topics.

**METHODS:** In 1992, the National Center for Health Statistics (NCHS) fielded the Youth Risk Behavior Survey (YRBS) as a follow-up to the National Health Interview Survey (NHIS). Because of concerns about accurate and complete reporting for the sensitive subjects included on the 1992 YRBS, the use of innovative methods to minimize risk of disclosure was explored. To ensure the privacy of the respondents so that participation and honesty could be maximized, the interviews were administered using a portable audiocassette tape player with headphones (PACTAPH), with the questions and answer categories recorded on tape. Interviewers also recorded information on the proximity of parents during the interview. Data from the 1992 National Household Survey on Drug Abuse (NHSDA), a survey in which the interviewer read the questions and the respondent marked an answer sheet, were used for comparison purposes. NHSDA and YRBS share similarities in sampling methodologies, can both be used to produce national estimates, and contain comparable questions relevant to the analysis.

**RESULTS/CONCLUSIONS:** This paper provides evidence consistent with findings of previous research of a relationship between the degree of privacy and the frequency of positive responses to sensitive questions by youths. The evidence suggests that the PACTAPH interviewing technique provides a greater degree of privacy from parental disclosure than do interviewer- and/or self-administered approaches. Youths clearly are concerned about parents or other household members learning of their responses and require the most secure setting and interview mode to respond honestly to sensitive questions. The use of the PACTAPH approach appears to provide the level of privacy necessary for maximum disclosure of sensitive behaviors. The findings of this research suggest that broad promises of anonymity or confidentiality are perhaps less important to honest reporting by youths than are assurances of privacy from the immediate threat of disclosure to parents.

Although most levels of reporting on sensitive behaviors appeared to increase slightly when parents were at home but out of the room, there were few significant differences within age groups, suggesting that the use of PACTAPH allayed respondent fears of disclosure. With one exception, estimates from the YRBS for three measures (smoking cigarettes in the past 30 days, ever trying marijuana and ever drinking alcohol) were higher than those from the NHSDA for three age groups (12–13, 14–15, 16–17). In addition, differences in estimates by parental proximity were larger in the NHSDA than on the YRBS, further suggesting support for the premise that the primary concern to youths is the immediate threat of disclosure to parents rather than anonymity or confidentiality.

## **The development and implementation of a new data collection instrument for the 1994 national household survey on drug abuse**

CITATION: Office of Applied Studies. (1996, April). *Development and implementation of a new data collection instrument for the 1994 National Household Survey on Drug Abuse* (HHS Publication No. SMA 96-3084, prepared under Contract No. 283-93-5409 for the Substance Abuse and Mental Health Services Administration). Rockville, MD: Substance Abuse and Mental Health Services Administration. [Prepared by L. A. Kroutil, L. L. Guess, M. B. Witt, L. E. Packer, J. V. Rachal, & J. C. Gfroerer]

PURPOSE/OVERVIEW: Since 1971, the National Household Survey on Drug Abuse (NHSDA) series has provided researchers and policymakers with key information about the extent of drug use and drug-related problems in the United States. Because of the importance of the survey, considerable attention has focused on ways to refine the NHSDA instrument to ensure accuracy of measurement. This report presents summaries of the background and history of the NHSDA, results of methodological studies on the NHSDA instrument, background on the implementation of the new questionnaire, and details of how the questionnaire was changed.

METHODS: The authors describe the background and rationale for the redesign of the NHSDA instrument, document the improvements made to the NHSDA instrument and the implementation plan for the new instrument, and assess the quality of the data obtained from the old and new versions of the 1994 instrument. To assess the quality of the data, the authors compare unweighted data from older versions of the 1994 NHSDA instrument with data from the new versions in terms of item nonresponse, inconsistent answers among related items, and overall nonresponse, including incomplete interviews. The authors also investigate whether differences that occur between estimates based on the old and new versions could be due to changes in the instrumentation, changes in the editing procedures, a combination of questionnaire and editing effects, the effect of statistical imputation procedures for handling missing data, or sampling error.

RESULTS/CONCLUSIONS: There was no indication that the new questionnaire had any effect on respondents' willingness to participate in the survey. The quality of data in the new instrument may be better than in the older version in many ways. The degree of missing data was lower in the new questionnaire. Also, inconsistencies between drug use recency and related variables were improved in the new instrument. Comparison of prevalence estimates from the old and new versions of the instrument for 1994 produced comparable estimates for most drugs and within most demographic subgroups. Although most differences were not statistically significant, some differences were. The higher rates of cigarette and smokeless tobacco use based on the new instrument may be related to differences in administration mode (self-administered in the new instrument vs. interviewer-administered in the old instrument). Some differences in prevalence estimates appeared to be due, at least in part, to differences in the format and content of the instruments.

## **Estimates of drug use prevalence in Miami from the 1991–1994 National Household Survey on Drug Abuse: Methodological report**

CITATION: Witt, M., Guess, L., Gfroerer, J., & Wright, D. (1996). *Estimates of drug use prevalence in Miami from the 1991–1994 National Household Survey on Drug Abuse: Methodological report*. Research Triangle Park, NC: Research Triangle Institute.

PURPOSE/OVERVIEW: From 1991 to 1993, the National Household Survey on Drug Abuse (NHSDA) sample was expanded to provide the ability to estimate drug use prevalence in six large metropolitan areas, including Miami. The resulting estimates for the Miami area indicated substantial decreases in illicit drug use from 1991 to 1993. These data were cited as evidence of the effectiveness of the Miami Coalition, an anti-substance abuse coalition formed in early 1990. However, anomalies seen in the data led the Substance Abuse and Mental Health Services Administration's (SAMHSA's) Office of Applied Studies (OAS) and the Research Triangle Institute (RTI), the contractor conducting the NHSDA, to investigate a number of factors that could have affected the NHSDA data during the 1991 to 1993 time period.

METHODS: The investigation involved a series of analyses of the NHSDA data, with a major emphasis on the impact of Hurricane Andrew, which hit Miami in August 1992. NHSDA data were analyzed separately for areas most affected by Hurricane Andrew and for the time periods before and after the hurricane.

RESULTS/CONCLUSIONS: The evidence suggests that the NHSDA-estimated decrease in illicit drug use was not due to any change in the design of the survey, in the estimation and editing methodology, or in the weighting procedures. There was considerable evidence to suggest that Hurricane Andrew did affect NHSDA results. Rates of drug use and perceived availability of drugs dropped after Hurricane Andrew. However, the demographic characteristics of the sample population did not seem to be affected by the hurricane. Additionally, although there were small decreases in NHSDA rates of current illicit drug use in Miami before Hurricane Andrew, these decreases generally were not statistically significant. Finally, the observed decrease in lifetime prevalence estimates made it questionable to conclude that there were real decreases in past month prevalence estimates.

# 1997

## **Repeated measures estimation of measurement bias for self-reported drug use with applications to the National Household Survey on Drug Abuse**

CITATION: Biemer, P. P., & Witt, M. (1997). Repeated measures estimation of measurement bias for self-reported drug use with applications to the National Household Survey on Drug Abuse. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (NIH Publication No. 97-4147, NIDA Research Monograph 167, pp. 439-476). Rockville, MD: National Institute on Drug Abuse.

PURPOSE/OVERVIEW: Direct estimates of response bias for self-reports of drug use in surveys require that essentially error-free determinations of drug use be obtained for a subsample of survey respondents. The difficulty of obtaining determinations that are accurate enough for estimating validity is well documented in the literature. Methods such as specimen (e.g., hair, urine) analysis, proxy reports, and the use of highly private and anonymous modes of interview all have to contend with error rates that may be only marginally lower than those of the parent survey. Thus, any methodology for direct validity estimation must rely to some extent on approximations and questionable assumptions.

METHODS: In this chapter, the authors considered a number of methods that rely solely on repeated measures data to assess response bias. Because the assumptions associated with these approaches do not require highly accurate second determinations, they may be more easily satisfied in practice. One such method for bias estimation for dichotomous variables that is considered in some detail provides estimates of misclassification probabilities in the initial measurement without requiring that the second measure be accurate or even better than the first. This methodology does require, however, that two subpopulations exist that have different rates of prevalence but whose probabilities of false-positive and false-negative error are the same. The applicability of these methods for self-reported drug use are described and illustrated using data from the National Household Survey on Drug Abuse (NHSDA).

RESULTS/CONCLUSIONS: In this chapter, a general model for studying misclassification in self-reported drug use was presented, and the model then was extended to the case where two measurements of the same characteristic are available for the sample of respondents. For the two-measurements case, the general model required seven parameters while only three degrees of freedom were available for estimation. Thus, some additional assumptions were required to reduce the set of unknown parameters to three or less. It was shown how the assumptions typically made for test-retest, true value, improved value, and Hui-Walter methods relate to the general model. Further, it was shown how the measures of reliability, measurement bias, estimator bias, mean squared error, false-negative probability, and false-positive probability can be defined in the context of the general model and how they may be estimated under the appropriate study designs.

Finally, the use of Hui and Walter's method for estimating misclassification error based on two erroneous reports was demonstrated. The reports may be self-reports, biological tests,

administrative record values, or any other measure. For the general case of two measurements, the Hui-Walter method used maximum likelihood estimation to obtain estimates of the false-negative and false-positive probabilities associated with each measurement, as well as the error-adjusted estimates of prevalence based on both measurements. The method required that the population be divided into two domains that have markedly different prevalence estimates and that satisfy the assumption of homogeneity of error probabilities.

## **Substance abuse in states and metropolitan areas: Model based estimates from the 1991–1993 National Household Surveys on Drug Abuse: Methodology report**

CITATION: Folsom, R. E., & Judkins, D. R. (1997). *Substance abuse in states and metropolitan areas: Model based estimates from the 1991–1993 National Household Surveys on Drug Abuse: Methodology report* (HHS Publication No. SMA 97-3140, Methodology Series M-1). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

PURPOSE/OVERVIEW: This report presents estimates of substance abuse for 26 States and 25 metropolitan statistical areas (MSAs). These estimates were developed from data collected in the National Household Survey on Drug Abuse (NHSDA) combined with local area indicators from a variety of sources. They were produced by the Substance Abuse and Mental Health Services Administration (SAMHSA) to provide State and local area policymakers with information on the prevalence of substance abuse behaviors and problems in their local areas. These estimates were an inexpensive alternative to the direct survey approach for describing substance abuse in State and local areas. They were based on a consistent methodology across areas and were constructed so that they summed to national estimates produced by the NHSDA.

METHODS: These State and MSA estimates were the result of a comprehensive small area estimation (SAE) project that included the development of an innovative methodology based on the methods used by other Federal agencies to meet needs for small area data. The methodology employed logistic regression models that combine NHSDA data with local area indicators, such as drug-related arrests, alcohol-related death rates, and block group–level characteristics from the 1990 census that were found to be associated with substance abuse. Work also was carried out to evaluate the model used to produce these estimates.

RESULTS/CONCLUSIONS: Considering a variety of evidence, the following conclusions were made: (1) The SAE model produced estimates of all measures that were much better than States could achieve by simply applying NHSDA national prevalence rates for demographic subgroups to the population distribution in their States. (2) A preponderance of evidence indicated that the estimates for alcohol, cigarette, and any illicit drug use were good in that they adequately reflected both levels of use and differences across States and MSAs in the level of use. (3) For past month use of any illicit drug but marijuana, past month use of cocaine, past year dependency on illicit drugs and dependency on alcohol, and need for treatment, the limited evidence indicated that these estimates also were reasonably good. (4) For arrests, past year treatment for illicit drug use, and past year treatment for alcohol, the quality assessments resulted in mixed findings, and these small area data may not have been good indicators of either differences between States or between MSAs, except in broad terms, or of the actual levels.

## Studies of nonresponse and measurement error in the National Household Survey on Drug Abuse

CITATION: Gfroerer, J., Lessler, J., & Parsley, T. (1997). Studies of nonresponse and measurement error in the National Household Survey on Drug Abuse. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (NIH Publication No. 97-4147, NIDA Research Monograph 167, pp. 273-295). Rockville, MD: National Institute on Drug Abuse.

PURPOSE/OVERVIEW: A summary of the results of a series of studies of nonresponse and measurement error in the National Household Survey on Drug Abuse (NHSDA) is given in this chapter. Two studies not previously reported, the Skip Pattern Experiment and the Census Match Study, are the primary focus of the chapter. The Skip Pattern Experiment involved a test of a modified NHSDA questionnaire that made extensive use of skip patterns in drug use questions. Compared with the standard NHSDA method, which avoided skip patterns, the modified questionnaire tended to produce lower rates of reported drug use. The Census Match Study involved linking 1990 NHSDA nonrespondent cases with data from the 1990 decennial census. Household and individual data for NHSDA nonrespondents were obtained from the census and used to characterize NHSDA nonresponse patterns in detail.

METHODS: A multilevel logistic model of response propensity identified the important predictors of nonresponse, including characteristics of the sampled person, the selected household, the neighborhood, and the interviewer.

RESULTS/CONCLUSIONS: Drug abuse surveys are particularly vulnerable to nonresponse and measurement error because of the difficulties in accessing heavy drug users and the likelihood that the illegal and stigmatized nature of drug abuse may lead to underreporting. The Skip Pattern Experiment confirmed once again that respondents' reporting of their drug use behavior was highly sensitive to the conditions under which they report. This conclusion made clear the need to proceed with great caution in interpreting differences in drug use rates obtained in different surveys. It also suggested caution in the implementation of new technologies, such as computer-assisted data collection, that undoubtedly will have some as yet unknown effect on respondents' willingness to report their drug use. The Skip Pattern Experiment may have implications in the introduction of these new technologies because one of the advantages of computer-assisted interviewing is the ease with which skips can be implemented. The Census Match Study demonstrated that response rates were not constant across various interviewer, respondent, household, and neighborhood characteristics. To the extent that rates of drug use varied by these same characteristics, bias due to nonresponse may be a problem. However, it was not always the case that low response rates occurred in conjunction with high drug use prevalence. Some populations with low response rates (e.g., older adults and high-income populations) tended to have low rates of drug use. On the other hand, some populations (e.g., residents of large metropolitan areas and men) had low response rates and high drug use rates. In estimating overall prevalence, many of these potential sources of bias would be in opposite directions and would therefore tend to cancel each other.

## **Prevalence of youth substance use: The impact of methodological differences between two national surveys**

CITATION: Gfroerer, J., Wright, D., & Kopstein, A. (1997). Prevalence of youth substance use: The impact of methodological differences between two national surveys. *Drug and Alcohol Dependence*, 47(1), 19-30.

PURPOSE/OVERVIEW: The purpose of this paper is to analyze the differences in the methodology used in two major federally sponsored surveys of drug use among youths.

METHODS: This study compares the prevalence rates yielded by the Monitoring the Future (MTF) study and the National Household Survey on Drug Abuse (NHSDA) by comparing the differences in response rates, precision, coverage, and data collection methods.

RESULTS/CONCLUSIONS: Although the MTF has a larger sample size of youths, an analysis of precision estimates reveals that its precision is similar to that of the NHSDA. After controlling for ages used, time of data collection, and dropouts, estimates of alcohol and drug use were significantly lower in the NHSDA than in the MTF. Although the exact cause of these differences could not be determined, the most likely reasons were the different interview environments (home vs. school), the questionnaires, and nonresponse bias in the MTF.

## **The validity of self-reported drug use data: The accuracy of responses on confidential self-administered answer sheets**

CITATION: Harrell, A. V. (1997). The validity of self-reported drug use data: The accuracy of responses on confidential self-administered answer sheets. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (NIH Publication No. 97-4147, NIDA Research Monograph 167, pp. 37-58). Rockville, MD: National Institute on Drug Abuse.

PURPOSE/OVERVIEW: Official records offer a relatively inexpensive, nonintrusive strategy for checking on the accuracy of self-reported drug use. The study reported here was designed as a criterion validity test of the National Household Survey on Drug Abuse (NHSDA) procedures. In criterion validity studies, two different measures of the same trait or experience are available: a candidate measure and an external, independent criterion measure that is treated as an error-free measure of the construct.

METHODS: In this study, the underreporting of illegal drug use was investigated in a sample of 67 former drug treatment clients by comparing their survey responses with clinic records on drug problems at time of admission. The criterion measures were based on self-reported marijuana, cocaine, hallucinogens, and heroin use. Drug treatment records were obtained from the files of publicly funded drug treatment programs in three States. The study followed the NHSDA interviewing and questionnaire procedures closely. To avoid bias from interviewer expectations and to protect the respondent's privacy, the sample of treatment clients was embedded in a larger sample of respondents. Interviewers were not told that the respondents had been treated for drug abuse. Special sample selection directions, tailored to match the target respondent's age and

gender, were used to select the former drug treatment client within the household, simulating the random selection screening instrument used in NHSDA. This analysis compared reports of past year and lifetime use of marijuana, cocaine, hallucinogens, and heroin by the former drug treatment clients with the drugs listed as problematic at time of admission to treatment. This analysis also examined factors that might have influenced the respondents' willingness or ability to respond accurately, such as the level of privacy during the interview and the amount of time between admission to the program and the interview.

**RESULTS/CONCLUSIONS:** The accuracy of reports compared with clinic records varied by drug, with the percentage of known users reporting their use highest for marijuana, followed by cocaine and hallucinogens, and lowest for heroin. Almost half of this sample of former treatment clients denied ever receiving drug treatment.

### **Recall decay and telescoping in self-reports of alcohol and marijuana use: Results from the National Household Survey on Drug Abuse (NHSDA)**

**CITATION:** Johnson, R. A., Gerstein, D. R., & Rasinski, K. A. (1997). Recall decay and telescoping in self-reports of alcohol and marijuana use: Results from the National Household Survey on Drug Abuse (NHSDA). In *Proceedings of the 1997 Joint Statistical Meetings, 52nd annual conference of the American Association for Public Opinion Research, Norfolk, VA* (pp. 964-969). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** To a large extent in the mid-1990s, the knowledge of life-cycle patterns of drug use in the United States was based on retrospective self-reports of survey respondents. Although most evidence for validity came from studies comparing self-reports of individuals dependent on narcotics with hospital and criminal justice records and with the results of urine tests (Nurco, 1985), such measures were too expensive to obtain in general population surveys. Instead, many general population surveys used the reinterview design to evaluate response error.

**METHODS:** The authors used a repeated cross-sectional design to evaluate response errors and analyzed changes in the distribution of responses of the same birth cohorts as measured in cross-sectional surveys conducted in different years. There were advantages and disadvantages to this method. The advantage is that pooling data from 10 National Household Surveys on Drug Abuse (NHSDAs) conducted between 1979 and 1995 allowed birth cohorts for a 16-year period at minimal expense; some disadvantage were the target population and methodological differences.

**RESULTS/CONCLUSIONS:** The authors used a cross-sectional design to show that NHSDA estimates of alcohol and marijuana incidence were biased downward by response error, especially at ages 10 to 14. The downward bias increased with the retention interval, making stable trend lines look increasing when estimated using a single survey. In addition, there was evidence that suggests forward telescoping, as well as some intentional concealment among older respondents.

## Is "up" right? The National Household Survey on Drug Abuse [review]

CITATION: Miller, P. V. (1997). Is "up" right? The National Household Survey on Drug Abuse [review]. *Public Opinion Quarterly*, 61(4), 627-641.

PURPOSE/OVERVIEW: For almost a decade, starting in the late 1980s, politicians quoted statistics from the National Household Survey on Drug Abuse (NHSDA) to claim that drug use among teenagers was rising. However, results from the 1996 NHSDA actually showed a decline in teenage drug use. This 1997 review discusses the NHSDA methodology to analyze the assumption that drug use was rising despite statistics to the contrary.

METHODS: A number of methodological factors, such as sampling, response rates, and data collection issues, are examined to assess their impact on prevalence estimates. The NHSDA sampling frame does not include individuals who are homeless (not living in shelters), nor does it include institutionalized individuals. There is no evidence on whether nonresponders to the NHSDA differ from responders in level of drug use. The NHSDA went through a number of methodological changes over the years aimed to improve survey quality and encourage reporting. These factors influence the likelihood of whether the NHSDA underreports drug use.

RESULTS/CONCLUSIONS: This examination of the NHSDA methodology suggests that the survey might have underestimated drug use. However, because no studies on the likelihood of overreporting had been conducted, it was impossible to say with conviction whether the estimates were an underreport. Validation studies do not answer all the questions surrounding data quality issues in the NHSDA; however, they do help to understand how different methods have an impact on survey estimates.

## The drug abuse treatment gap: Recent estimates

CITATION: Woodward, A., Epstein, J., Gfroerer, J., Melnick, D., Thoreson, R., & Willson, D. (1997). The drug abuse treatment gap: Recent estimates. *Health Care Financing Review*, 18(3), 5-17.

PURPOSE/OVERVIEW: In the mid-1990s, it was acknowledged that the National Household Survey on Drug Abuse (NHSDA) underestimated the prevalence of drug use in the United States due to coverage error and social desirability bias. In this article, the authors attempted to correct for the underreporting by applying new estimation procedures.

METHODS: To control for sampling errors and underreporting due to social desirability in the NHSDA, three smaller data sources were used to assist with the ratio estimation. The three data sources were (1) the National Drug and Alcoholism Treatment Unit Survey (NDATUS), later known as the Uniform Facility Data Set (UFDS); (2) the Drug Services Research Survey (DSRS); and (3) the Uniform Crime Report (UCR). Ratio estimation relies on the assumption that estimates can be improved when a known value exists for a related variable. To assess the effects on the need for treatment for drug abuse, the term "need" had to be redefined.

**RESULTS/CONCLUSIONS:** The new definition for "need" included both a clinical and epidemiological perspective, making the definition more comprehensive and reliable for use in reporting estimates. The ratio estimation procedures improved the NHSDA estimates for the number of people in need of treatment because it corrected for coverage error and underreporting.

### **Ratio estimation of hardcore drug use**

**CITATION:** Wright, D., Gfroerer, J., & Epstein, J. (1997). Ratio estimation of hardcore drug use. *Journal of Official Statistics*, 13(4), 401-416.

**PURPOSE/OVERVIEW:** In the mid-1990s, the prevalence of hard-core drug use was consistently underestimated in national surveys. The goal of this paper was to improve the estimates for hard-core drug use from the National Household Survey on Drug Abuse (NHSDA) by using additional sources and applying ratio estimation.

**METHODS:** Ratio estimation improves an estimate by comparing it with known totals for a related variable. To estimate hard-core drug use, the authors made use of two known populations related to drug use. First was the number of people in treatment centers for drug abuse, which was taken from the National Drug and Alcoholism Treatment Unit Survey (NDATUS). The second known population was the number of arrests during the past year taken from the Federal Bureau of Investigation's Uniform Crime Reports (UCRs). To get estimates of hard-core drug users, the NHSDA ratio of drug users to people in treatment was multiplied by the known total population of people in treatment. The same procedure was applied to the number of arrests during the past year. These methods yielded two estimates for the number of hard-core drug users. Another step of ratio estimation was to use the two known populations together to create one single, more accurate estimate.

**RESULTS/CONCLUSIONS:** The standard errors for the ratio estimation were similar to the standard errors for the simple expansion estimator that was alternatively used. The ratio estimates were considered an improvement over other estimation techniques. However, they still did not account for underreporting in people who were not arrested or in treatment centers. An advantage to the estimates provided by ratio estimation is that it could provide more than just the "bottom line" and could yield estimates for such subgroups as race, region, gender, and income.

### **The use of external data sources and ratio estimation to improve estimates of hardcore drug use from the NHSDA**

**CITATION:** Wright, D., Gfroerer, J., & Epstein, J. (1997). The use of external data sources and ratio estimation to improve estimates of hardcore drug use from the NHSDA. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (NIH Publication No. 97-4147, NIDA Research Monograph 167, pp. 477-497). Rockville, MD: National Institute on Drug Abuse.

**PURPOSE/OVERVIEW:** In the mid-1990s, levels of hard-core drug use were especially difficult to estimate because of the relative rarity of the behavior, the difficulty of locating hard-core drug users, and the tendency to underreport stigmatized behavior. This chapter presents a new application of ratio estimation, combining sample data from the National Household Survey on Drug Abuse (NHSDA) together with population counts of the number of individuals arrested in the past year from the Uniform Crime Report (UCR) and the number of those in drug treatment programs in the past year from the National Drug and Alcoholism Treatment Unit Survey (NDATUS). The population counts served as a benchmark accounting for undercoverage and underreporting of hard-core drug users.

**METHODS:** In this discussion, the focus was on the ratio estimate's ability to reduce bias (in particular, the undercounting of hard-core drug users in the NHSDA) given a true population value of a related variable. To make the discussion more concrete, the estimation procedure was applied to four separate, but overlapping, measures of hard-core drug use for 1992: the number of past year users of heroin, weekly users of cocaine in the past year, past year users who were more dependent on some illicit drug, and past year intravenous drug users.

**RESULTS/CONCLUSIONS:** Overall statements about ratio estimation for the NHSDA include the following: (1) ratio estimation did not fully account for underreporting in the NHSDA; (2) because ratio estimation could be looked at as an adjustment to the NHSDA analytic weights (which were based on a probability sample design), it provided analytic capabilities that were not possible in other methods; (3) the ratio estimation model, as applied in this case, relied primarily on regularly updated and consistently collected data from the NHSDA, NDATUS, and UCR, and a relatively small number of easily understood assumptions; and (4) because ratio estimation relied primarily on the NHSDA sample design and weighting, it was possible to develop estimates of the variances of ratio-adjusted estimates.

# 1998

## **Major design changes in the National Household Survey on Drug Abuse**

CITATION: Barker, P., Gfroerer, J., Caspar, R., & Lessler, J. (1998). Major design changes in the National Household Survey on Drug Abuse. In *Proceedings of the 1998 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Dallas, TX* (pp. 732-737). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: This paper reports on interim results of methodological research carried out to test design changes eventually introduced in the 1999 National Household Survey on Drug Abuse (NHSDA). The main focus is on a feasibility test conducted in 1996 that compared administering the survey through paper-and-pencil interviewing (PAPI) and computer-assisted interviewing (CAI). The paper also briefly describes laboratory testing of 12-month frequency of use questions and plans to increase the NHSDA sample sizes in 1999 to produce State-level estimates.

METHODS: In the 1996 CAI Feasibility Experiment, 400 interviews were carried out using either PAPI or CAI. For the CAI version, core items on substance use were administered using audio computer-assisted self-interviewing (ACASI). The CAI version consisted of two versions, one that incorporated skip patterns into the survey instrument and another that did not and thus, "mirrored" the PAPI version. Interviewer observation questions also were asked.

RESULTS/CONCLUSIONS: The main result of the 1996 CAI Feasibility Experiment was that a CAI approach to gathering NHSDA data was quite feasible and had a number of benefits to offer over the PAPI approach, including increased reporting of past year and past month marijuana and cocaine use and increased perceptions of privacy. It also was found that the skip pattern version of CAI was shorter than the CAI version that "mirrored" PAPI by about 10 minutes. The cognitive laboratory testing of the 12-month frequency of use items revealed that these questions were difficult to answer due to a long recall period and that the response categories combined total number of days with periodicity estimates. An alternative response format was developed that allowed the respondent to select the unit for reporting days of use.

## **Reports of smoking in a national survey: Data from screening and detailed interviews, and from self- and interviewer-administered questions**

CITATION: Brittingham, A., Tourangeau, R., & Kay, W. (1998). Reports of smoking in a national survey: Data from screening and detailed interviews, and from self- and interviewer-administered questions. *Reports About Smoking: Annals of Epidemiology*, 8(6), 393-401.

PURPOSE/OVERVIEW: Using data from the 1994 National Household Survey on Drug Abuse (NHSDA), the authors tested the primary hypothesis that using self-administered questionnaires versus interviewer-administered questionnaires would increase reports of cigarette smoking

among adolescents. The study also compared respondent and proxy responses during a brief screening interview with the responses collected during a more detailed interview later.

**METHODS:** Approximately 22,000 respondents aged 12 or older were interviewed using a national area probability sample. During the screening interview, basic demographic information and current smoking status were collected for all household members. Using this information, one or more respondents from the household were selected to complete the main interview. Sometimes the respondent who completed the screening interview also was selected for the main interview. Sample members were randomly assigned to receive either interviewer-administered smoking questions or self-administered smoking questions. Logistic regression was used to compare the rates of smoking by interview mode and differences in reporting of smoking between the screening and the main interview.

**RESULTS/CONCLUSIONS:** Although the self-administered questions showed higher reports of smoking than interviewer-administered questions, the differences were not significant for the overall population. When the analyses were restricted to adolescents, there was a marginally significant main effect for mode of interview, with self-administered questions yielding higher reporting. The proportion of current smokers reported in the screening interview was significantly lower than in the main interview. Discrepancies between the screening interview and the main interview occurred 4 times more often when the screening respondent was different from the interview respondent. This supports the notion that the screening respondent was a proxy and was not necessarily cognizant of the smoking status of the rest of the household. When the same respondent was interviewed in the screener and in the main interview, the highest rate of discrepancy occurred in the adolescent age group, supporting the hypothesis that smoking is sensitive for adolescents but not adults. In addition, asking multiple questions about smoking leads to higher reports of smoking, particularly in respondents who may be reluctant to admit their smoking status.

### **Testing ACASI procedures to reduce inconsistencies in respondent reports in the NHSDA: Results from the 1997 experimental field test**

**CITATION:** Caspar, R. A., Lessler, J. T., & Penne, M. A. (1998). Testing ACASI procedures to reduce inconsistencies in respondent reports in the NHSDA: Results from the 1997 experimental field test. In *Proceedings of the 1998 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Dallas, TX* (pp. 750-755). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** One of the most significant potential benefits of converting the National Household Survey on Drug Abuse (NHSDA) to a computer-assisted format is the opportunity to resolve inconsistent data at the time of the interview rather than editing the data to deal with inconsistencies after the fact. However, maintaining the privacy benefits of the audio computer-assisted self-interviewing (ACASI) component of the interview requires that the respondent be able to resolve inconsistencies for many items on his or her own. Thus, one of the goals of the NHSDA conversion work was to develop a method for resolving inconsistent data that the respondent could easily understand and complete without significant intervention by the interviewer.

**METHODS:** Based on the authors' own hypotheses about how inconsistencies should be identified and resolved, they developed a resolution methodology that combined two components. First, at the verify stage, respondents are asked whether an answer they have entered is in fact correct. So, for example, when a 20-year-old respondent reports that she was 51 the first time she drank alcohol (a clearly inconsistent answer), the computer was programmed to verify that this information was correct. If the respondent indicated that the information was incorrect, she was routed back to answer the question again (perhaps this time entering the age of her first drink as 15). A second component incorporates the resolution of seemingly inconsistent answers. For example, a respondent who indicated drinking alcohol on 15 days in the past 12 months, but then reported drinking alcohol on 25 days in the past 30 days, first would be asked to verify the last entry keyed. If the respondent indicated that the entry was correct, then he or she was routed to a question that identified the inconsistency and provided the respondent with an opportunity to correct one or both of the entries.

**RESULTS/CONCLUSIONS:** Based on the results reported, the authors feel that the inconsistency resolution methodology employed in the 1997 CAI Field Experiment was successful. The methodology improved the consistency of the data collected without adversely affecting respondent cooperation or burden. Using this methodology in future implementations of the NHSDA will allow the Substance Abuse and Mental Health Services Administration (SAMHSA) to capitalize on the numerous benefits of the ACASI technology while minimizing one of the potential pitfalls—that respondent errors and inconsistencies are not identified and corrected at the time of interview. As of 1998, work is under way to incorporate inconsistency resolution screens into the 1999 computer-assisted personal interviewing (CAPI) and ACASI NHSDA instrument. The authors anticipate that future research in this area will be conducted to determine whether respondents can resolve inconsistencies between items that are not closely spaced in the interview. If this proves to be possible, the authors feel that future NHSDA instruments may incorporate inconsistency resolution screens of this type as well.

### **Estimating substance abuse treatment need from a national household survey**

**CITATION:** Epstein, J. F., & Gfroerer, J. C. (1998, May). Estimating substance abuse treatment need from a national household survey. In J. Epstein (Ed.), *Analyses of substance abuse and treatment need issues* (HHS Publication No. SMA 98-3227, Analytic Series A-7, pp. 113-125). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** For purposes of planning future demand on the health care system in general and the substance abuse treatment system in particular, it is important to be able to develop estimates of the number of people needing treatment for substance abuse (i.e., treatment need) on a regular basis. In order to do this, the measurement of treatment need must distinguish low-intensity substance use from drug use that requires intervention. The overall prevalence of substance use is a poor absolute measure of problem substance use. Estimating treatment need is a difficult problem. Drug and alcohol consumption patterns and their consequences are complicated and dynamic. The modalities and philosophies of treatment are diverse. The applicability of even well-ensconced and tested diagnostic criteria must be reestablished as new drugs and ways of administering them appear. Measuring treatment need involves both a

scientific and clinical understanding of the substance use problem. Estimating how many people in the general population need treatment is a different problem from diagnosing the need for treatment for an individual based on history taking, physical examination, and information in previous records. In most household surveys, it would be impractical to perform physical examinations or take a detailed history. In addition, household surveys may not fully cover key populations that may have a significant number of people needing treatment, such as incarcerated and homeless individuals. The main focus of this paper describes attempts in the 1990s to estimate treatment need based on the National Household Survey on Drug Abuse (NHSDA). The NHSDA is potentially a valuable source for estimating treatment need because it is the only ongoing large national survey of substance use covering most of the population. Thus, if a reasonable method for estimating treatment need from the NHSDA could be developed, it would provide more timely estimates than other sources and could be used to measure changes in treatment need over time.

**METHODS:** To estimate the effects of changes in the questionnaire and variable construction, both the new and old versions of the survey were fielded in 1994. The "old" version of the survey, which was actually the same as the 1993 questionnaire, is referred to as 1994-A, while the "new" version of the survey is referred to as 1994-B.

**RESULTS/CONCLUSIONS:** Comparisons between the new questionnaire (1994-B) and the old questionnaire (1994-A) showed that estimates of the rate of lifetime illicit drug use from the new questionnaire were approximately 8 percent less than estimates from the old questionnaire. Estimates of the rate of past year illicit drug use were approximately 12 percent less with the new questionnaire. On the other hand, estimates of the rate of past month illicit drug use were approximately 5 percent higher with the new questionnaire. The new questionnaire produced approximately 29 percent higher estimates of the rate of weekly cocaine use and approximately 6 percent higher estimates of the rate of daily marijuana use.

## **Nonresponse in household interview surveys**

**CITATION:** Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York City, NY: Wiley.

**PURPOSE/OVERVIEW:** The 1990 National Household Survey on Drug Abuse (NHSDA) was one of six large Federal or federally sponsored surveys used in the compilation of a dataset that then was matched to the 1990 decennial census for analyzing the correlates of nonresponse. In addition, data from surveys of NHSDA interviewers were combined with those from these other surveys to examine the effects of interviewer characteristics on nonresponse.

**METHODS:** Information from the NHSDA, such as interviewer notes, segment listing sheets, and segment maps, as well as other data sources were used to match sampled housing units from the 1990 NHSDA with housing units from the 1990 decennial census. Variables on household characteristics from the 1990 census and area characteristics were used as predictors of cooperation on the NHSDA and the other surveys in the dataset. For characteristics of interviewers, a survey was administered to NHSDA interviewers during training. Among the items asked about were (1) education; (2) primary reason for their interviewing job and the

existence of another paid job; (3) experience, including years worked and number of organizations worked for in the last 5 years; and (4) questions regarding interviewer attitudes and expectations. These variables were used as predictors of cooperation on the NHSDA as well as the other surveys.

**RESULTS/CONCLUSIONS:** Overall, 97.2 percent of NHSDA sample housing units (a total of 4,619 units) were successfully matched to the decennial census, with 97.8 percent of interviewed housing units ( $n = 1,407$ ) and 96.7 percent of nonresponse housing units ( $n = 2,681$ ) successfully matched. A total of 280 NHSDA interviewers completed the survey of interviewer characteristics. The main findings were as follows: (1) Once contacted, those with lower socioeconomic status were no less likely to cooperate than those with higher socioeconomic status; there was instead a tendency for those in high-cost housing to refuse survey requests, which was partially accounted for by residence in high-density urban areas. (2) The tendency for individuals in military service and in non-English-speaking households to cooperate appeared to be a function of household composition. (3) Households with children or young adults were more likely to cooperate; single-adult households were less likely to cooperate. (4) After controlling for the effects of household size, households with elderly residents tended to cooperate. (5) The tendency of racial and ethnic minorities to participate was accounted for by lower socioeconomic status. (6) Densely populated, high-crime urban areas had lower cooperation rates, which could be accounted for by other ecological variables, as well as different household compositions in urban versus rural areas. (7) Interviewers with more interviewing experience tended to achieve higher cooperation rates than those with less experience, but it was not clear if this was due to higher attrition among less successful interviewers or additional skills gained by experienced interviewers. (8) There was some evidence that interviewers with higher levels of confidence in their ability to gain participation achieved higher cooperation rates.

### **Adjusting survey estimates for response bias: An application to trends in alcohol and marijuana use**

**CITATION:** Johnson, R. A., Gerstein, D. R., & Rasinski, K. A. (1998). Adjusting survey estimates for response bias: An application to trends in alcohol and marijuana use. *Public Opinion Quarterly*, 62(3), 354-377.

**PURPOSE/OVERVIEW:** Lifetime prevalence estimates and trends in drug use are primarily based on self-reports. Validity testing usually compares the reports given by special populations with administrative records and biochemical tests. However, these types of validity measures are not possible in a general population study; therefore, most studies rely on test-retest procedures.

**METHODS:** Instead of the traditional reinterview method for assessing validity, this study used a repeated cross-sectional design. This method utilized birth cohorts and examined changes in the distribution of their responses over individual cross-sectional surveys. Using data from 9 years of the National Household Survey on Drug Abuse (NHSDA) from 1982 to 1995, trends in birth cohorts were evaluated over 13 years. Changes in survey design in the NHSDA over the 13 years may have introduced bias in assessing response error.

**RESULTS/CONCLUSIONS:** Following birth cohorts using a repeated cross-sectional design is cost-effective in assessing self-report error in studies of drug and alcohol use. Using an exponential decay model of response bias revealed lower reporting of drug use as the time between the drug use and survey response increased. This bias distorts trends in alcohol and drug use, making stable drug use trends misleadingly appear increasing. Further analysis revealed that respondents tended to forward telescope the age for first alcohol use. In addition, intentional nondisclosure of marijuana use at a young age increased as the respondent increases in age, which led to underreporting of lifetime prevalence rates for marijuana use.

### **Development of computer assisted interviewing procedures for the National Household Survey on Drug Abuse (NHSDA): Design and operation of the 1997 CAI field experiment**

**CITATION:** Lessler, J. T., Witt, M., & Caspar, R. (1998). Development of computer assisted interviewing procedures for the National Household Survey on Drug Abuse (NHSDA): Design and operation of the 1997 CAI field experiment. In *Proceedings of the 1998 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Dallas, TX* (pp. 738-743). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** A large-scale field experiment that examined the use of computer-assisted interviewing (CAI) for the National Household Survey on Drug Abuse (NHSDA) was conducted in the last quarter of 1997. The design of the 1997 field experiment was based on the results of a 1996 CAI feasibility experiment and subsequent cognitive laboratory testing, power calculations, and discussions as to the operational feasibility of various designs. The authors compared alternative versions of the audio computer-assisted self-interviewing (ACASI) portion of the CAI interview in a factorial design.

**METHODS:** The authors compared these alternatives with each other and with the results from the methodology employing a combination of a paper-and-pencil interviewing (PAPI) and self-administered answer sheets. They conducted the experiment in the fourth quarter of 1997 and used the Quarter 4 1997 NHSDA survey results as a comparison group.

**RESULTS/CONCLUSIONS:** The screening and interview response rates in the 1997 experimental field test were lower than those achieved in the main study. The overall screening response rates was 86.8 percent, which was about 7 percent lower than that achieved in similar areas in the national NHSDA. About 2.5 percent of this shortfall was due to the failure to obtain access to restricted housing; 3.5 percent was due to increased refusals. It is unlikely that the electronic screener contributed to the failure to obtain access to restricted housing. However, the authors were not able, from this study alone, to verify that using the Newton screener did not contribute to increased refusals to the screening.

## **Discussion of issues facing the NHSDA**

**CITATION:** Miller, P. V. (1998). Discussion of issues facing the NHSDA. In *Proceedings of the 1998 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Dallas, TX* (p. 762). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** The papers in this session report on the results of a field test that compared paper-and-pencil questionnaires with several versions of a computer-assisted interviewing (CAI) measurement of drug use. The study discussed is one of many studies conducted to assess the effects of converting the National Household Survey on Drug Abuse (NHSDA) from paper-and-pencil interviewing (PAPI) to CAI.

**METHODS:** N/A.

**RESULTS/CONCLUSIONS:** The results support a priori views that CAI has the potential to improve data quality in NHSDA. Findings include indications that a streamlined CAI version of the questionnaire can produce higher reporting of drug use and that inconsistency checking built into the CAI instrument can improve reporting. The authors also noted the success of ACASI technology in providing respondent privacy, handling complex skip patterns, and resolving inconsistencies between responses.

## **Effects of experimental audio computer-assisted self-interviewing (ACASI) procedures on reported drug use in the NHSDA: Results from the 1997 CAI field experiment**

**CITATION:** Penne, M. A., Lessler, J. T., Bieler, G., & Caspar, R. (1998). Effects of experimental audio computer-assisted self-interviewing (ACASI) procedures on reported drug use in the NHSDA: Results from the 1997 CAI field experiment. In *Proceedings of the 1998 Joint Statistical Meetings, American Statistical Association, Social Statistics Section, Dallas, TX* (pp. 744-749). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** The 1997 computer-assisted interviewing (CAI) field experiment evaluated the impact on reported drug use of using alternative versions of an audio computer-assisted self-interviewing (ACASI) version of the National Household Survey on Drug Abuse (NHSDA).

**METHODS:** Alternative versions of the ACASI questionnaire were examined using a factorial design conducted during the fourth quarter of 1997. A subsample of the Quarter 4 national NHSDA, which used a combination of a paper-and-pencil interviewing (PAPI) version and self-administered questionnaire (SAQ) answer sheets, comprised the control group for the study. A fuller description of the design can be found in Lessler, Witt, and Caspar (1998) in this volume. In this paper, the authors examined reported drug use and compared the experimental ACASI factors to each other and to the control group. They also presented information on overall differences between ACASI and the control group.

**RESULTS/CONCLUSIONS:** The prevalence data indicate that ACASI will yield higher estimates of drug use, particularly for youths. The debriefing data indicate that this is due to the privacy-enhancing features of the method.

### **Drug abuse treatment: Data limitations affect the accuracy of national and state estimates of need**

**CITATION:** U.S. General Accounting Office. (1998). *Drug abuse treatment: Data limitations affect the accuracy of national and state estimates of need* (GAO/HEHS-98-229, Report to Congressional Requesters, Health, Education, and Human Service Division). Washington, DC: U.S. Government Printing Office.

**PURPOSE/OVERVIEW:** The Federal Government annually provides approximately \$3 billion for drug abuse prevention and treatment activities. However, determining the need for treatment services for the general population, as well as for specific subpopulations, may be problematic due to limitations in national and State data on treatment need. This report describes the Substance Abuse and Mental Health Services Administration's (SAMHSA) efforts to estimate drug abuse treatment need on a national basis and obtain State estimates of drug abuse treatment need.

**METHODS:** The authors interviewed and obtained documents from officials in SAMHSA's Center for Substance Abuse Treatment (CSAT), the Office of Applied Studies (OAS), and the Office of the Administrator. The authors also held discussions with experts in the substance abuse research community. They reviewed needs assessment information submitted by States. In addition, the authors attended a CSAT-sponsored workshop that included all States with current State Treatment Needs Assessment Program (STNAP) contracts in which States reported on their needs assessment studies.

**RESULTS/CONCLUSIONS:** Reliable assessments of treatment need are an essential component to accurately target treatment services. Although SAMHSA is improving its national estimates through the expansion of the National Household Survey on Drug Abuse (NHSDA) sample, the survey is still likely to underestimate treatment need. Also, STNAP's goals to help States develop estimates of treatment need and improve State reporting of need data have not been fully accomplished. Even though States are required to provide estimates of treatment need as part of their block grant applications, not all States report this information, and some of the data reported are inaccurate. SAMHSA recognizes the need to increase State reporting and has set a target for increasing the number of States that provide the information. It also recognizes that the overall quality of the data reported is problematic.

### **A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use**

**CITATION:** Wright, D. L., Aquilino, W. S., & Supple, A. J. (1998). A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. *Public Opinion Quarterly*, 62(3), 331-353.

**PURPOSE/OVERVIEW:** This article assesses the impact of using computer-assisted self-interviewing (CASI) to collect self-report data on sensitive questions about drugs, alcohol, and mental health. The study also measures the interaction that computerization and respondent characteristics have on the survey responses.

**METHODS:** Using a national multistage area probability sample, 3,169 interviews were conducted with respondents aged 12 to 34. Approximately two thirds completed the computer self-administered questionnaire (SAQ), and the other one third completed a paper-and-pencil interviewing (PAPI) SAQ. The interviews used the drug and mental distress questions from the 1995 National Household Survey on Drug Abuse (NHSDA) questionnaire. Ordinary least squares regression was used to measure the effect of mode on prevalence estimates.

**RESULTS/CONCLUSIONS:** For all variables except for tobacco use, the computer SAQ produced higher estimates than the paper SAQ for adolescents. The greatest effects were seen on the most sensitive items. This supports the notion that adolescents are more familiar with and have a better understanding of computers than adults, which makes them more comfortable using the computer SAQ. There was a significant mode by mistrust interaction in older respondents for alcohol and drug use. Respondents who were mistrustful had higher estimates of drug and alcohol use in the PAPI SAQ than in the CASI SAQ.

## **Hierarchical models applied to the National Household Survey on Drug Abuse (NHSDA)**

**CITATION:** Wright, D., & Zhang, Z. (1998). Hierarchical models applied to the National Household Survey on Drug Abuse (NHSDA). In *Proceedings of the 1998 Joint Statistical Meetings, American Statistical Association, Survey Methods Section, Dallas, TX* (pp. 756-761). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** A fairly comprehensive discussion of the methodology in hierarchical modeling was introduced in *Hierarchical Linear Models: Applications and Data Analysis Methods* (1992) by Anthony S. Bryk and Stephen W. Raudenbush. A significant part of that book focused on hierarchical structures in the field of education and the necessity of using an estimation methodology that reflects the structure of the data. A natural question is how that methodology might apply to the drug use field. Much of the drug use analysis research over the years has utilized simple logistic regression to estimate relationships between drug use and a variety of person-level variables. The authors wanted to explore how hierarchical models could be applied to the National Household Survey on Drug Abuse (NHSDA) data and what the consequences would be of ignoring the hierarchy. Since its inception, the NHSDA has been a multiphase stratified sample of primary sampling units (counties or groups of counties), segments (blocks or block groups), households, and individuals. The result of this complex nested design is that using traditional techniques of variance estimation that ignore the clustering of sample cases and the use of sample weights tends to overstate the significance of many findings. A number of statistical packages have been developed using Taylor series methods or replication methods, such as SUDAAN<sup>®</sup> (RTI) and WESVAR<sup>®</sup> (WESTAT), that take this structure into account. There was no widely used software for hierarchical mixed models, however, until hierarchical linear modeling (HLM).

**METHODS:** This paper explored some of the special circumstances in applying hierarchical models to NHSDA. The authors decided to use the data from six oversampled cities from the 1991 to 1993 NHSDAs. With a sample of approximately 2,500 individuals per city for each of the 3 years, there was a total sample of about 45,000 people. Each of the six cities (Washington, DC; Chicago; Miami; New York City; Denver; and Los Angeles) was selected with certainty, and within each city, segments were sampled at the first stage. Within segment, usually about 10 to 12 individuals aged 12 or older were selected. If the authors wanted to make inference to the collection of the six large cities as typical large cities, then they could consider four levels of hierarchy: city, segment, family, and person. Two issues arose immediately: Were there sufficient number of observations at each hierarchical level to support the analysis, and was it necessary to include all four levels of hierarchy in the estimation? By the latter, it is meant, "Was there sufficient variability at each stage to be necessary for inclusion?"

**RESULTS/CONCLUSIONS:** In the analysis, a significant portion of the variation (about 80 percent or more) was attributed to person-level variation. The person-level variation really included the variation between families as well. Because family-level variables, such as family drug use, parental attitudes about drug use, and family management practices and family conflict, had been identified as risk factors for adolescent drug use, it was possible that a significant amount of what had been labeled as person-level variation was really between-family variation. These components of variance needed to be estimated. The issue of small sample sizes at the family level in particular needed to be explored. In the 1997 and later NHSDAs, it was attempted to collect data on a random sample of pairs of individuals within a sample of households. Using Goldstein's formula for the impact of this vis-a-vis ordinary least squares (OLS), the  $(n-1)$  would be equal to 2, and the inflation factor above and beyond simple random sampling would be  $(1+\rho)$ . For a large  $\rho$ , this might make a considerable difference. If the level-2 sample sizes were larger, then even a small  $\rho$  could make a difference. So, it was important to understand that any conclusions with respect to the use of HLM versus OLS applied only to the NHSDA and only with the current NHSDA sample sizes. The impact on other datasets or on the NHSDA with different sample sizes could be quite different. Although the variables that had been analyzed were interesting, there was a need to consider other drug-related variable scales and to extend the analysis to dichotomous data, such as past year use of marijuana or any illicit drug.

# 1999

## **Usability testing for self-administered survey instruments: Conversion of the National Household Survey on Drug Abuse as a case study**

CITATION: Caspar, R., & Barker, P. (1999). Usability testing for self-administered survey instruments: Conversion of the National Household Survey on Drug Abuse as a case study. In R. Banks, C. Christie, J. Currall, J. Francis, P. Harris, B. Lee, J. Martin, C. Payne, & A. Westlake (Eds.), *ASC '99: Leading survey & statistical computing into the new millennium. Proceedings of the 3rd Association for Survey Computing International Conference* (pp. 79-89). Chesham, Buckinghamshire, United Kingdom: Association for Survey Computing.

PURPOSE/OVERVIEW: Computer-assisted interviewing (CAI) has prompted many changes in the way survey instruments are designed and administered. For example, computer-assisted instruments can contain significantly more complex routings than those developed for paper administration. At the same time, however, computer-assisted instruments must be designed so that they can be easily navigated by the end user. When, as in the case of self-administered questionnaires, the end user is a survey respondent who may have little or no experience with computers and may not be particularly motivated to complete the survey task, it is especially important that the computer-assisted instrument be designed to be both easily understood and easily completed. This paper describes the usability testing that has been completed as part of the work to convert the National Household Survey on Drug Abuse (NHSDA) from a paper-and-pencil interviewing (PAPI) instrument to a CAI instrument with an extensive audio computer-assisted self-interviewing (ACASI) component.

METHODS: Forty subjects were recruited to take part in the testing. Vignettes were used to simulate response inconsistencies in order to reduce costs. Respondents first were asked to verify their last response, with responsibility for the inconsistency not explicitly placed on the respondent. Original responses were displayed on screen, and respondents were asked to identify the incorrect response. The study did not address the issue of why respondents give inconsistent responses or how they would feel to have their inconsistencies identified by the computer.

RESULTS/CONCLUSIONS: An estimated 28 percent of respondents triggered a check, 45 percent of these respondents triggered a true inconsistency, and youth respondents accounted for 60 percent of the true inconsistencies triggered. Respondents triggered an average of 1.2 true inconsistencies and 1.2 "verify only." Eighty-one percent of true inconsistencies were resolved. There was a very low rate of "Don't Know" or "Refused" responses, no significant increase in interview length, and no significant increase in breakoff rates. Respondents were slightly more likely to report that they gave accurate and complete answers. Respondents also were slightly more likely to report difficulty using the computer, and interruptions and distractions were the most common reason for inconsistent data based on interviewer reports. In conclusion, usability testing improved the NHSDA instrument fielded in 1999. The approach to developing usable ACASI instruments was not so different from developing usable self-administered PAPI questionnaires. Usability testing requires planning—the researcher must allocate additional time during the development phase.

## Population coverage in the National Household Survey on Drug Abuse

CITATION: Chromy, J. R., Bowman, K. R., Crump, C. J., Packer, L. E., & Penne, M. A. (1999). Population coverage in the National Household Survey on Drug Abuse. In *Proceedings of the 1999 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Baltimore, MD* (pp. 576-580). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: This paper presents a brief description and history of the National Household Survey on Drug Abuse (NHSDA), a description of the household screening process, definitions of the coverage terms reported in this paper, a summary of overall measures across recent years, and some results by selected domains.

METHODS: For the purposes of this paper, the authors take a broad view of survey coverage in household screening surveys. They include weighted person response rates, weighted household response rates, and a measure of frame coverage based on recent experience from the NHSDA.

RESULTS/CONCLUSIONS: NHSDAs provide a large sample base for studying overall coverage rates, including those associated with both nonresponse error and frame coverage error. Results presented on frame coverage have been based on comparisons of nonresponse-adjusted estimates to external data sources. Because both are subject to error, some of the frame coverage rates need to be interpreted with caution. In particular, the Hispanic population coverage is shown to be higher than other groups, but that may be an artifact of the intercensal projections for Hispanics rather than a true measure of frame coverage. Screening response, interview response, and frame coverage have been discussed separately as if they were independent processes. In fact, this is not the case. Poor screening response can have differential effects on frame coverage if the nonscreened dwelling units tend to be ones that have a higher proportion of a certain domain (e.g., younger individuals or blacks). This confounding of the screening coverage and frame coverage occurs because the development of dwelling unit rosters within dwelling units is actually the final stage of frame development. This confounding of effects may provide further support for studying all forms of undercoverage—screening nonresponse, interview nonresponse, and frame noncoverage—as an overall phenomenon.

## You want 75,000 unique maps by WHEN?

CITATION: Curry, R. (1999). You want 75,000 unique maps by WHEN? In *Proceedings of the Environmental Systems Research Institute (ESRI) User Conference, San Diego, CA, July 26-30, 1999*. Redlands, CA: Environmental Systems Research Institute (ESRI).

PURPOSE/OVERVIEW: Research Triangle Institute (RTI) has been contracted to run and manage the National Household Survey on Drug Abuse (NHSDA) for the Substance Abuse and Mental Health Services Administration (SAMHSA). The 1999 NHSDA was the largest survey RTI had managed up to that date. Because of the size of the survey and the requirement for large numbers of accurate maps in a relatively short period of time, geographic information system (GIS) technology became an important part of the household survey process. The Survey Automated Mapping System (SAMS) and Setting and Zooming (SAZ) were developed to satisfy the mapping requirements of the survey. A total of 7,200 sample segments were chosen from

every State in the country. Over 1,200 field interviewers (FIs) needed to know where to administer the survey in those 7,200 geographic locations.

**METHODS:** For each segment, a set of maps was generated showing county, tract, segment, and block boundaries to help survey takers collect data from the correct locations. Over a 10-week period, an automated mapping system generated over 75,000 of these unique maps for the FIs.

**RESULTS/CONCLUSIONS:** By using a high-powered UNIX workstation, ARC/INFO software, an online TIGER/Line database, along with data files and driver files, one researcher was able to use SAMS to produce high volumes of maps in a relatively short period of time while minimizing the number of errors and reruns. Other survey staff interacted with the mapping information and assisted in creating the data files for SAMS as well as producing zoom maps when needed. The field survey staff used these maps as a means to locate the survey area and to assist in the interview process.

### **Validity of drug use reporting in a high-risk community sample: A comparison of cocaine and heroin survey reports with hair tests**

**CITATION:** Fendrich, M., Johnson, T. P., Sudman, S., Wislar, J. S., & Spiehler, V. (1999). Validity of drug use reporting in a high-risk community sample: A comparison of cocaine and heroin survey reports with hair tests. *American Journal of Epidemiology*, 149(10), 955-962.

**PURPOSE/OVERVIEW:** Toxicological tests often performed on hair specimens are the standard for assessing the accuracy of self-reported drug use data. However, most studies of this sort use specialized populations, such as inmates or people in drug use treatment programs. The purpose of this study was to use toxicological results to assess the accuracy of self-reports in a community sample.

**METHODS:** Using the procedures and instrument from the 1995 National Household Survey on Drug Abuse (NHSDA), respondents were assigned randomly into one of two conditions. The first condition was the control group and replicated the NHSDA procedures almost entirely. The second condition was the experimental group, which utilized a cognitive interviewing procedure. Respondents in both groups were not offered an incentive for cooperating in the survey. However, after respondents completed the survey, they were offered a \$10 incentive for a specimen of their hair. Hair samples were given by 322 respondents.

**RESULTS/CONCLUSIONS:** The results demonstrated that the toxicological tests revealed higher prevalence estimates for drug use than the survey measures. Respondents were more likely to report heroin use than cocaine use. In addition, hard-core users who had the most concentration of cocaine in their hair were more likely to report drug use than others. Respondents also were more inclined to report lifetime use than recent use, indicating that lifetime self-report estimates may be more reliable than recent-use self-reports. The findings of this study are limited by the possibility that the \$10 incentive disproportionately motivated drug users to participate.

## Measuring interstate variations in drug problems

CITATION: McAuliffe, W. E., LaBrie, R., Lomuto, N., Betjemann, R., & Fournier, E. (1999). Measuring interstate variations in drug problems. *Drug and Alcohol Dependence*, 53(2), 125-145.

PURPOSE/OVERVIEW: This paper describes the process used to create the Drug Problem Index (DPI), a measure that combined a collection of drug abuse indicators. The DPI was designed to compare severity of drug abuse and dependence problems across States.

METHODS: All measures included in the index came from empirical and theoretical evidence of connections to drug abuse and incorporated information from the National Drug and Alcoholism Treatment Unit Survey (NDATUS), the National Association of State Alcohol and Drug Abuse Directors (NASADAD), the Federal Bureau of Investigation's (FBI) Uniform Crime Reporting (UCR) system, and the National Household Survey on Drug Abuse (NHSDA). Measures were combined into an index by summing their *z*-scores using nominal weights. Measures for drug abuse severity had to be tested for reliability and validity to be included in the index. The measures were tested for both construct validity and external validity. Comparisons of the DPI with other measures of drug abuse also were conducted.

RESULTS/CONCLUSIONS: Three components for the DPI were selected: drug-coded mortality, drug-defined arrest, and drug-treatment client rates. Results of the reliability analysis revealed that the index demonstrated stable estimates from year to year. The index also demonstrated convergent and construct validity. In addition, the DPI correlated with the block grant drug need index. However, it did not correlate with other measures of drug abuse from NHSDA. Neither the NDATUS drug-treatment-only client rate nor the NASADAD drug admissions rate had a significant correlation with either the NHSDA direct or model-based estimates of the past year drug treatment rate.

## 2000

### **Use of alternating logistic regression in studies of drug-use clustering**

CITATION: Bobashev, G. V., & Anthony, J. C. (2000). Use of alternating logistic regression in studies of drug-use clustering. *Substance Use & Misuse*, 35(6-8), 1051-1073.

PURPOSE/OVERVIEW: This paper aims to provide readers with a detailed review of the alternating logistic regression (ALR) methodology as a tool for studying clustering and community-level drug involvement.

METHODS: The authors applied the ALR analysis method to a publicly available national dataset from the National Household Survey on Drug Abuse (NHSDA) in order to examine the factors of clustered drug involvement in the United States. The variables of interest were measures of marijuana use on a weekly or more frequent basis. The ALR method can take into account sociodemographic characteristics providing estimates for different subgroups.

RESULTS/CONCLUSIONS: Clustering of drug use was apparent using the ALR methodology. Clustering was most notable in daily drug use and slightly less notable with weekly drug use. Sociodemographic characteristics of the individual, such as age, gender, income, and race, did not affect the estimates for marijuana use. However the authors' approach was unable to find strong determinants or factors that caused the clustering. They concluded that an analysis with a dataset that provided more indicators on the economic and social conditions of the neighborhood would be useful in further understanding the causes of clustering in drug use.

### **Experience with the generalized exponential model (GEM) for weight calibration for NHSDA**

CITATION: Chen, P., Penne, M. A., & Singh, A. C. (2000). Experience with the generalized exponential model (GEM) for weight calibration for NHSDA. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 604-607). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: The National Household Survey on Drug Abuse (NHSDA) is designed to estimate prevalence of both licit and illicit drug use in the United States for various demographic and geographic domains. Since 1999, it has become a statewide survey that includes 50 States and the District of Columbia. The target population includes civilian, noninstitutionalized individuals aged 12 or older. Eight States (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas), referred to as the "big" States, have a sample designed to yield 3,600 respondents per State, while the remaining 43 "small" States have a sample designed to yield 900 respondents per State. The total sample size is 66,706 individuals (corresponding to 51,821 dwelling units [DUs] selected at the second phase out of 169,166 DUs screened at the first phase) with a low of 756 for Nevada to a high of 1,280 for Utah among

"small" States, and a low of 2,669 for New York and a high of 4,681 for California among "big" States.

**METHODS:** The 1999 NHSDA data for the East South Central Census Division is used to illustrate results obtained by fitting the generalized exponential model (GEM) at both the first phase for nonresponse (nr) and poststratification (ps) of the screener DUs and the second phase for ps for selected individuals followed by nr and ps for respondents.

**RESULTS/CONCLUSIONS:** The summary of characteristics of various models fitted is given, as are summary statistics in terms of the unequal weighting effect (UWE), extreme values (ev), and outwisor proportions, as well as distributional characteristics of the weight distribution. It is seen that percent ev is reasonable after ps; therefore, there is no need for an extra ev step. Point estimates and standard errors (SEs) are compared for baseline and final models across a set of drug use variables. For confidentiality reasons, only ratios of point estimates (final over the baseline) are presented; however, individual relative standard errors (RSEs) (SE over the estimate) are shown. Two types of RSE are given; one is unadjusted signifying no adjustment for calibration, and the other is ps-adjusted RSE denoting a sandwich-type formula to adjust for ps, as given in Vaish, Gordek, and Singh (2000) in this volume. It is seen that the two (baseline and final) RSEs are quite comparable. Interestingly, the final RSE can be lower than the baseline RSE although it has more covariates. In cases where the final is higher, it is only marginally so, showing no problems of overfitting.

### **The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification**

**CITATION:** Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 598-603). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** There exist methods in the literature that impose bounds on the adjustment factor for poststratification (ps). See, for example, Deville and Särndal (1992), Rao and Singh (1997), and the review by Singh and Mohl (1996). However, they do not directly restrict the adjusted weight from being too extreme. In this paper, the authors consider the problem of developing a unified approach of weight calibration to address the above three concerns such that there are built-in controls on the adjustment factors to prevent the adjusted weight from being too extreme.

**METHODS:** For this purpose, the logit-type model of Deville and Särndal (1992), denoted by DS in the sequel, is generalized to allow for more general and unit-specific bounds. A review of the DS model is provided in Section 2, and the proposed model is described in Section 3. The asymptotic properties of the proposed calibration estimator are presented in Section 4, and a comparison with alternative methods is given in Section 5. Finally, Section 6 contains numerical results comparing different methods using the 1999 National Household Survey on Drug Abuse (NHSDA) data followed by concluding remarks in Section 7.

**RESULTS/CONCLUSIONS:** Unlike earlier methods, the generalized exponential model (GEM) provides a unified calibration tool for weight adjustments for extreme values (ev), nonresponse (nr), and ps. Of special interest is its capability to have a built-in control on ev. Under suitable superpopulation modeling and assuming that the bounds on the adjustment factors are prespecified, the resulting calibration estimators were shown to be asymptotically consistent with respect to the distribution for nr and ps, and derivation of the asymptotic Taylor variance estimation formulas analogous to the ones based on residuals for the regression estimator (used for ps) was outlined. In the authors' experience, GEM is a useful practical alternative to the methods of raking-ratio and Deville-Särndal while providing comparable results.

### **Developing computer-assisted interviewing (CAI) for the National Household Survey on Drug Abuse**

**CITATION:** Lessler, J. T., Caspar, R. A., Penne, M. A., & Barker, P. R. (2000). Developing computer-assisted interviewing (CAI) for the National Household Survey on Drug Abuse. *Journal of Drug Issues*, 30(1), 9-34.

**PURPOSE/OVERVIEW:** To ensure the privacy of respondents' answers to the National Household Survey on Drug Abuse (NHSDA), the sponsor of the survey in the late 1990s planned to convert it to computer-assisted interviewing (CAI) with an audio computer-assisted self-interviewing (ACASI) section for the most sensitive items. As a result of the survey being self-administered, respondents needed to be able to reconcile inconsistent data during the interview without the assistance of an interviewer.

**METHODS:** Based on prior findings and research, an ACASI instrument was created that included skip instructions and allowed respondents to reconcile inconsistent answers. A  $2 \times 2 \times 2$  factorial design was used. The first factor explored whether multigate questions would yield higher prevalence estimates than single-gate questions. The second factor was the ability of respondents to complete data quality consistency checks and its effect on the length of the survey. The third and final factor explored whether multiple opportunities to report drug use would yield higher estimates. The experiment was embedded in the last quarter of the 1997 NHSDA sample. There were approximately 1,100 respondents in the experiment group and 3,500 in the control group.

**RESULTS/CONCLUSIONS:** The results showed small but significant time increases for all three factors: multigate questions, inconsistency checks, and multiple opportunities to report drug use. However, the total extra time for these factors was only 1 minute, which was not large enough to rule out elimination on time alone. Contrary to expectations, multigate questions were not found to lead to higher prevalence estimates. Respondents were able to resolve inconsistency checks without unduly increasing the length of the interview. The multiple questions on drug use did not consistently increase or decrease prevalence rates.

## **Design consistent small area estimates using Gibbs algorithm for logistic models**

CITATION: Shah, B. V., Barnwell, B. G., Folsom, R., & Vaish, A. (2000). Design consistent small area estimates using Gibbs algorithm for logistic models. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 105-111). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: This paper describes the software solutions used to produce age-specific small area estimates and associated pseudo-Bayes posterior intervals for the 50 States and the District of Columbia for the National Household Survey on Drug Abuse (NHSDA).

METHODS: Survey weights were incorporated in conditional posterior distributions to achieve design-consistent asymptotic mean vector and covariance matrices. Proper inverse Wishart priors were used for the covariance matrices. The bias of the estimated fixed effects and covariance components was examined for a three-stage cluster sample design. The software consists of two procedures, PROC GIBBS and PROC GSTAT. PROC GIBBS is used to simulate the posterior distribution for fixed and random effects models using the Gibbs algorithm. PROC GSTAT estimates the prevalence rate of a substance for each block group for each type of individual within the block group for each simulated cycle. Weighted averages of the block group estimates are used to produce county- or State-level estimates.

RESULTS/CONCLUSIONS: The software was able to fit four age-group-specific models simultaneously for over 70,000 observations, 30 to 40 predictors in each model totaling over 100 fixed effects and over 350 random effects. Over 10,000 Gibbs cycles were completed in 10 to 12 hours on a personal computer using the Win95 operating system (300 MHz Pentium II with 256MB RAM). MLwiN and BUGS, two software packages designed to estimate multilevel models, were unable to complete a single cycle. For States with large sample sizes, PROC GIBBS and PROC GSTAT produced estimates that were close to design-based estimates using SUDAAN<sup>®</sup>.

## **Bias corrected estimating function approach for variance estimation adjusted for poststratification**

CITATION: Singh, A. C., & Folsom, R. E., Jr. (2000). Bias corrected estimating function approach for variance estimation adjusted for poststratification. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 610-615). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: In this paper, the authors approached the problem of variance estimation adjusted for coverage bias (via poststratification [ps]) by using the conceptual similarity to the variance estimation adjusted for nonresponse (nr) bias, results for which are well known. To this end, it was observed that use of calibration equations for nr and ps adjustments led quite naturally to the nonoptimal estimating function (EF) framework introduced by Binder (1983) for variance estimation for finite population parameters under a quasi-design based framework (i.e., under the joint superpopulation model and design-based distribution), and the

optimal EF framework introduced by Godambe (1960) and Godambe and Thompson (1989) for finite or infinite population parameters. In view of the superpopulation model required for correcting coverage bias or nr or both, the authors therefore proposed a bias-corrected estimating function (BCEF) approach under a quasi-design based framework for estimating variance adjusted for ps or nr or both. The proposed BCEF method was motivated by observing the similarity between ps and nr when one takes the perspective of coverage bias reduction in ps. The BCEF method is based on a simple semiparametric model built on estimating functions that are commonly used for estimating parameters for modeling nr and ps adjustment factor.

**METHODS:** The authors considered the problem of finding a Taylor linearization variance estimator of the poststratified or the general calibration estimator such that four goals, somewhat analogous to those of Sarndal, Swensson, and Wretman (1989), were met. These goals were (1) the variance estimator should be consistent under a joint design-model distribution (i.e., in a quasi-design based framework), (2) the variance estimator should have a single form with general applicability, (3) the model postulated for the quasi-design based framework should be driven by the real need for unbiased point estimation, and (4) the model should be expected to have sensible conditional properties under a conditional inference outlook whenever it can be suitably defined.

**RESULTS/CONCLUSIONS:** Using the property inherent to estimating functions, the BCEF method provided a sandwich variance estimate adjusted for ps or nr, or both (this is simply the inverse of the Godambe information matrix), which has a simple yet general form useful for computer automation. Also, using the Godambe finite sample optimality criterion of estimating functions, it was shown that the point and variance estimators (whenever the solution of the estimating function exists) were unique analogous to maximum likelihood estimation for parametric models.

## **Variance estimation adjusted for weight calibration via the generalized exponential model with application to the National Household Survey on Drug Abuse**

**CITATION:** Vaish, A. K., Gordek, H., & Singh, A. C. (2000). Variance estimation adjusted for weight calibration via the generalized exponential model with application to the National Household Survey on Drug Abuse. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 616-621). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** In this paper, the authors studied the impact of weight adjustment factors for nonresponse (nr) and poststratification (ps) on the variance of the calibrated sample estimate when adjustment factors were modeled via the generalized exponential model (GEM) of Folsom and Singh (2000) with suitable predictors and bounding restrictions on the adjustment factors.

**METHODS:** Using the bias-corrected estimating function (BCEF) approach of Singh and Folsom (2000), the estimation problem was cast in the form of estimating equations, which in turn was linearized to obtain a sandwich-type variance estimate of the calibrated estimator.

The method was applied to the 1999 National Household Survey on Drug Abuse (NHSDA), and numerical results comparing variance estimates with and without adjustments were presented.

**RESULTS/CONCLUSIONS:** The main points are as follows: (1) although the H matrix looks complicated, it followed a nice pattern and could easily be calculated and inverted by using SAS IML; (2) even with a large number of predictor variables used in calibration, the variance estimates adjusted for calibration seemed remarkably stable and, in general, showed gains in efficiency after calibration; and (3) the BCEF methodology was very general and could easily be adapted to other types of calibration techniques. The authors planned to do a validation study by computing resampling variance estimates using Jackknife and comparing the results with the BCEF approach based on the Taylor method. It may be noted that the Jackknife method for obtaining adjusted variance could be quite tedious for large datasets and with somewhat elaborate nr/ps models.

# 2001

## Discussion notes: Session 5

CITATION: Camburn, D., & Hughes, A. (2001). Discussion notes: Session 5. In M. L. Cynamon & R. A. Kulka (Eds.), *Seventh Conference on Health Survey Research Methods* (HHS Publication No. PHS 01-1013, pp. 251-253). Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

PURPOSE/OVERVIEW: The importance of data collection at the State and local level is recognized as a requirement for surveillance and monitoring systems. The presentations in the conference session this paper discusses describe Federal surveillance systems that include as a goal providing State and local area estimates.

METHODS: The session this paper discusses addressed issues affecting the systematic collection of State- and local-level data. The four surveys reported on in the session include the Behavioral Risk Factor Surveillance System (BRFSS), the National Household Survey on Drug Abuse (NHSDA), the National Immunization Survey (NIS), and the State and Local Area Integrated Telephone Survey (SLAITS). The topics discussed include (1) variation in data quality; (2) timeliness of data release; (3) balancing national, State, and local needs; (4) within-State estimates; and (5) analyzing and reporting.

RESULTS/CONCLUSIONS: (1) Concerns about variation in data quality centered on variation in State response rates, cultural differences in the level of respondent cooperation, interviewer effects, and house effects. (2) Survey producers are doing all they can to release the appropriate data in a timely manner, to document limitations, and to minimize microdata disclosure risk. (3) Standardized, well-controlled methodologies may be an advantage in some circumstances, but not others. State-level control over content is important, particularly for within-State analysis, although States are interested in comparability across States. (4) One limitation of current State-level surveillance systems is that resource and time constraints restrict the amount of data that can be collected within individual States for smaller geographic domains or for demographic subgroups. (5) An important issue for surveillance systems collecting data and providing estimates that cover a large number of areas is determining appropriate analysis methods and identifying appropriate methods for reporting the data that include area-specific data quality indicators. For example, in NHSDA, direct estimates are provided for the eight largest States while model-based estimates are calculated for the remaining States. Eventually, NHSDA plans to provide direct estimates for all 50 States.

## **Variance models applicable to the NHSDA**

**CITATION:** Chromy, J., & Myers, L. (2001). Variance models applicable to the NHSDA. In *Proceedings of the 2001 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** When planning or modifying surveys, it is often necessary to project the impact of design changes on the variance of survey estimates. Initial survey planning requires evaluation of the efficiency of alternative sample designs. Efficient designs meet all major requirements at the minimum cost and can only be evaluated in terms of appropriate variance and cost models. This paper focuses on a fairly simple variance model that nevertheless accounts for binomial variation, stratification, intracluster correlation effects, variable cluster sizes, differential sampling of age groups and geographic areas, and residual unequal weighting.

**METHODS:** The authors examined data from the 1999 National Household Survey on Drug Abuse national sample to develop variance model parameter estimates. They then compared modeled variance estimates with direct estimates of variance based on the application of design-based variance estimation using the SUDAAN<sup>®</sup> software.

**RESULTS/CONCLUSIONS:** The modeled relative standard errors provided a realistic approximation to those obtained from design-based estimates. Because they expressed the variance in terms of design parameters, they were useful for evaluating the impact of alternative designs configurations. The simple (unweighted) approach to variance component estimation appeared to provide useful results in spite of ignoring the weights. The impact of unequal weighting was treated as a multiplicative factor. Although the unequal weighting effect can be controlled to some extent by the survey design, the impact of nonresponse and weight adjustment for nonresponse and for calibration against external data can only be controlled in a general way. The unequal weighting effects were not easily subject to any optimization strategy. The authors concluded that the model treatment of variable cluster sizes, particularly for small domains, should be useful in developing variance models for a wide variety of applications.

## **Estimation for person-pair drug-related characteristics in the presence of pair multiplicities and extreme sampling weights for NHSDA**

**CITATION:** Chromy, J. R., & Singh, A. C. (2001). Estimation for person-pair drug-related characteristics in the presence of pair multiplicities and extreme sampling weights for NHSDA. In *Proceedings of the 2001 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** In the National Household Survey on Drug Abuse (NHSDA), Brewer's method is adapted for selecting 0, 1, or 2 individuals for the drug survey from the screened dwelling unit (DU) selected at the first phase. Typically, the parameter of interest is at the person level and not the pair level (e.g., at the parent level in the parent-child data). However, pair data are used for estimation because the study variable is measured only through a pair. In estimation,

two major problems arise. One is that of multiplicities because for a given domain, several pairs in a household could be associated with the person, and the multiplicities being domain specific make it difficult to produce a single set of calibrated weights. The other is that of extreme weights due to the possibility of small pair selection probabilities depending on the age groups, which might lead to unstable estimates.

**METHODS:** For the first problem, the authors propose to do the above calibration simultaneously for controls for key domains. For the second problem, they propose a Hajek-type modification, which entails calibration to controls obtained from larger samples from previous phases. Extreme weights are further reduced by a repeat calibration under bound restrictions while continuing to meet controls.

**RESULTS/CONCLUSIONS:** It is clear that weights based on pairwise probabilities are required for many drug behavior analyses of the NHSDA data. For this purpose, the analyst needs to make some fundamental decisions about defining population parameters when the person has the same relationship (parent of or child of) to more than one person in the household. For the two problems of multiplicities and extreme weights that might arise in pair data analysis, it was shown how the estimator could be adjusted in the presence of multiplicities and how the weights could be calibrated to alleviate the problem of extreme weights.

### **Substance use survey data collection methodologies and selected papers [Commentary]**

**CITATION:** Colliver, J., & Hughes, A. (2001). Substance use survey data collection methodologies and selected papers [Commentary]. *Journal of Drug Issues*, 31(3), 717-720.

**PURPOSE/OVERVIEW:** This paper presents a commentary on the methodological differences between three national surveys that study substance use among adolescents and young adults: the Monitoring the Future (MTF) study, the National Household Survey on Drug Abuse (NHSDA), and the Youth Risk Behavior Survey (YRBS).

**METHODS:** This paper reviews the current literature discussing the differences in estimates of drug use for the three surveys as a result of differences in documentation, sampling, and survey design.

**RESULTS/CONCLUSIONS:** The comparative studies sponsored by the Office of Assistant Secretary for Planning and Evaluation provide an excellent resource for understanding the methodological differences in the MTF, the NHSDA, and the YRBS that contribute to the discrepancy in estimates of drug use provided each year.

### **Coverage, sample design, and weighting in three federal surveys**

**CITATION:** Cowan, C. D. (2001). Coverage, sample design, and weighting in three federal surveys. *Journal of Drug Issues*, 31(3), 599-614.

**PURPOSE/OVERVIEW:** This paper compares and contrasts coverage and sampling used in three national surveys on drug use among teenagers: the Monitoring the Future (MTF) study, the National Household Survey on Drug Abuse (NHSDA), and the Youth Risk Behavior Survey (YRBS).

**METHODS:** This review starts by comparing the national teenage population coverage of each of the three surveys to assess the effects of coverage error. Next, the sample design for each study is compared, as well as changes to the methodology made over the years. Finally, the weighting procedures used are analyzed.

**RESULTS/CONCLUSIONS:** The author concluded that all three studies were well designed, and it was difficult to make recommendations to any of these surveys for improving the assessment of drug abuse among teenagers because it could affect other key variables or subgroups in the individual surveys that are not being compared. However, the author indicated that one recommendation that may not negatively affect the validity of the studies was an in-depth coverage study to assess both frame coverage and nonresponse.

### **Mode effects in self-reported mental health data**

**CITATION:** Epstein, J. F., Barker, P. R., & Kroutil, L. A. (2001). Mode effects in self-reported mental health data. *Public Opinion Quarterly*, 65(4), 529-549.

**PURPOSE/OVERVIEW:** This article measures the mode effect differences between audio computer-assisted self-interviewing (ACASI) and an interviewer-administered paper-and-pencil interview (I-PAPI) on respondent reports of mental health issues. Four mental health modules on major depressive episode, generalized anxiety disorder, panic attack, and agoraphobia were taken from the World Health Organization's Composite International Diagnostic Interview (CIDI) Short Form.

**METHODS:** The ACASI data were collected in a large-scale field experiment on alternative ACASI versions of the National Household Survey on Drug Abuse (NHSDA). The field experiment was conducted from October through December 1997. The comparison group was comprised of a subsample of the Quarter 4 1997 NHSDA. In the field experiment, mental health questions were administered by ACASI to 865 adults. In the comparison group, mental health questions were administered to a sample of 2,126 adults using I-PAPI. Logistic regression models were used to assess the differences in reporting mental health syndromes by mode of administration. Estimates were made overall and for several demographic variables, such as age, race/ethnicity, gender, education level, geographic region, and population density subgroups while controlling for confounding variables and interactions.

**RESULTS/CONCLUSIONS:** For most measures, the percentages of people reporting a mental health syndrome were higher for ACASI than I-PAPI. Overall differences were significant only for major depressive episode and generalized anxiety disorder. This study suggests that respondents report higher levels of sensitive behavior with ACASI than with I-PAPI likely due to a perception of greater privacy with ACASI.

## **Impact of computerized screenings on selection probabilities and response rates in the 1999 NHSDA**

CITATION: Eyerman, J., Odom, D., Chromy, J., & Gfroerer, J. (2001). Impact of computerized screenings on selection probabilities and response rates in the 1999 NHSDA. In *Proceedings of the 2001 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: The National Household Survey on Drug Abuse (NHSDA) is an ongoing Federal Government survey that tracks substance use in the United States with face-to-face interviews in a national probability sample. In the past, a paper screening instrument was used by field interviewers to identify and select eligible households in the sample frame. The paper screening instrument was replaced with a computerized instrument in 1999. The computerized instrument standardized the screening process and reduced the amount of sampling procedural error in the survey. This paper identifies a type of procedural error that is possible when using paper screening instruments and evaluates its presence in the NHSDA before and after the transition to a computerized screening instrument.

METHODS: The impact of the error is examined against response rates and substance use prevalence estimates for 1999.

RESULTS/CONCLUSIONS: Results suggest that paper screening instruments are vulnerable to sampling procedural error and that the transition to a computerized screening instrument may reduce the amount of the error.

## **Examining prevalence differences in three national surveys of youth: Impact of consent procedures, mode, and editing rules**

CITATION: Fendrich, M., & Johnson, T. P. (2001). Examining prevalence differences in three national surveys of youth: Impact of consent procedures, mode, and editing rules. *Journal of Drug Issues*, 31(3), 615-642.

PURPOSE/OVERVIEW: This review compares contact methods and mode used in three national surveys on drug use conducted in 1997: the Monitoring the Future (MTF) study, the National Household Survey on Drug Abuse (NHSDA), and the Youth Risk Behavior Survey (YRBS).

METHODS: Differences in information presented during the informed consent process are compared for all three studies to evaluate its impact on prevalence estimates of drug abuse. Next, the mode for the three studies was compared focusing on where the study took place (school vs. home environment), when the study took place, and how it was conducted (self-administered vs. interviewer-administered and paper-and-pencil interviewing [PAPI] vs. computer-assisted interviewing [CAI]). Finally, data-editing procedures for inconsistent responses and missing responses were analyzed for each survey.

**RESULTS/CONCLUSIONS:** Comparisons of these three surveys suggested that the consent process and mode used in the 1997 NHSDA may have contributed to the lower prevalence estimates compared with the other 1997 studies. The NHSDA consent process required more parental involvement and presented more consent information than did the process used by the other two studies, which may have inhibited respondent reporting. Differences in editing procedures did not appear to account for any differences in prevalence estimates for the three surveys.

### **Learning from experience: Estimating teen use of alcohol, cigarettes, and marijuana from three survey protocols**

**CITATION:** Fowler, F. J., Jr., & Stringfellow, V. L. (2001). Learning from experience: Estimating teen use of alcohol, cigarettes, and marijuana from three survey protocols. *Journal of Drug Issues*, 31(3), 643-664.

**PURPOSE/OVERVIEW:** This paper compares prevalence estimates for drug use among teenagers in the three national surveys funded by the Federal Government: the Monitoring the Future (MTF) study, the National Household Survey on Drug Abuse (NHSDA), and the Youth Risk Behavior Survey (YRBS).

**METHODS:** Because the three surveys rely on different modes for data collection, comparisons had to be made across similar groups. For that reason, rates at which adolescents in grades 10 and 12 reported drug use were compared. In addition, comparisons were made between males and females and across racial/ethnic groups, such as whites, Hispanics, and blacks. Finally, trends in drug use from 1993 to 1997 were compared across surveys.

**RESULTS/CONCLUSIONS:** Comparisons across these surveys were hard to draw because any differences in prevalence estimates were due to numerous confounding factors associated with coverage, sampling, mode, questionnaires, and data collection policies. A solution suggested is for each of these studies to set aside a small amount of money that can be used for collaborative studies of comparisons of the methods used.

### **Substance use survey data collection methodologies and selected papers [Commentary]**

**CITATION:** Gfroerer, J. (2001). Substance use survey data collection methodologies and selected papers [Commentary]. *Journal of Drug Issues*, 31(3), 721-724.

**PURPOSE/OVERVIEW:** The Office of the Assistant Secretary for Planning and Evaluation sponsored five comparative papers to assess differences in methodologies used by three national surveys on drug use among adolescents.

**METHODS:** The papers compared the National Household Survey on Drug Abuse (NHSDA), the Monitoring the Future (MTF) study, and the Youth Risk Behavior Survey (YRBS).

Differences in sampling, mode, and other data collection properties were assessed to determine their impact on the differences in prevalence estimates from the three surveys.

**RESULTS/CONCLUSIONS:** Finding differences in estimates can provide information to inform an understanding of the current estimates. The findings of these papers also will help target specific areas for more methodological research. Although these studies were sponsored to explain differences in estimates between the surveys, it also is important to note several of the consistencies in these surveys, such as in demographic differences and trends over time.

### **State estimates of substance abuse prevalence: Redesign of the National Household Survey on Drug Abuse (NHSDA)**

**CITATION:** Gfroerer, J., Wright, D., & Barker, P. (2001). State estimates of substance abuse prevalence: Redesign of the National Household Survey on Drug Abuse (NHSDA). In M. L. Cynamon & R. A. Kulka (Eds.), *Seventh Conference on Health Survey Research Methods* (HHS Publication No. PHS 01-1013, pp. 227-232). Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

**PURPOSE/OVERVIEW:** Starting with the 1999 survey, the National Household Survey on Drug Abuse (NHSDA) sample was expanded and redesigned to improve its capability to estimate substance use prevalence in all 50 States. This paper discusses the new NHSDA design, data generated from the new design, and issues related to its implementation. Also provided are a summary of the old survey design and a discussion of other major changes implemented in 1999, such as the conversion of the survey from paper-and-pencil interviewing (PAPI) to computer-assisted interviewing (CAI).

**METHODS:** The pre-1999 survey design and the new survey design are compared and discussed in terms of various methodological implications: (1) data collection methodology, (2) questionnaire content, (3) data limitations, (4) sample design, (5) issues in implementing the expanded sample, (6) estimates that the new design will provide, and (7) estimation and analysis issues.

**RESULTS/CONCLUSIONS:** The findings include the following: (1) Although the basic research methodology has remained unchanged, the conversion from PAPI to CAI incorporated features designed to decrease burden and increase data quality, including electronic screening, range edits, and consistency checks throughout the questionnaire, and inconsistency resolution in audio computer-assisted self-interviewing (ACASI). (2) Although the content of the 1999 CAI questionnaire is similar to the 1998 PAPI questionnaire, the CAI interview length is considerably shorter than the PAPI interview length. (3) Although the methods used in the survey have been shown effective in reducing reporting bias, there is still probably some unknown level of underreporting that occurs. (4) In response to a need for comparable State-level estimates of substance abuse prevalence, in 1999 SAMHSA expanded the survey sample. This was determined to be feasible based on prior experiences with modeling for selected States, as well as a sampling plan for 1999 that would facilitate State-level estimation. (5) The size and distribution of the 1999 sample units across the 50 States posed a challenge for data collection

operations. In spite of extensive training, the inexperience of new interviewers hired for the 1999 survey expansion led to a decline in response rates relative to prior NHSDAs. (6) The sample was designed to produce both model-based and sample-based State-level estimates of a variety of substance use measures. (7) Estimation and analysis issues include comparability of NHSDA State estimates, comparisons of NHSDA State-level estimates to other surveys in States, assessing trends within States, and data release and disclosure limitation.

## **Understanding the differences in youth drug prevalence rates produced by the MTF, NHSDA, and YRBS studies**

CITATION: Harrison, L. D. (2001). Understanding the differences in youth drug prevalence rates produced by the MTF, NHSDA, and YRBS studies. *Journal of Drug Issues*, 31(3), 665-694.

PURPOSE/OVERVIEW: This paper explores methodological differences in three national surveys of drug use in order to explain the discrepancy in prevalence estimates between the three sources.

METHODS: The three surveys examined in this paper are the National Household Survey on Drug Abuse (NHSDA), the Monitoring the Future (MTF) study, and the Youth Risk Behavior Survey (YRBS). This paper explores the validity of the survey estimates provided by these studies by comparing their differences in methodology, such as survey design, anonymity, confidentiality, and question context.

RESULTS/CONCLUSIONS: Despite the purpose of this paper being to explore the differences in estimates in these three surveys, analyses revealed that the estimates might be more similar than originally thought. The confidence intervals for several variables in the surveys overlapped, and trend analyses in drug use for the surveys followed the same pattern. Any differences found between the survey estimates were likely a result of many minor methodological differences between the surveys. No one survey can be shown to be more accurate than another, and in fact using several different sources provides a more informed overall picture of drug use among youths. The author concluded that each survey should continue using its current methodology because although different, each fulfills a particular need. However, more studies could be done to continue learning about the impact of methodological differences in each survey.

## **Substance use survey data collection methodologies: Introduction**

CITATION: Hennessy, K. H., & Ginsberg, C. (Eds.). (2001). Substance use survey data collection methodologies: Introduction. *Journal of Drug Issues*, 31(3), 595-598.

PURPOSE/OVERVIEW: There are three annual surveys funded by the U.S. Department of Health and Human Services that collect data on the prevalence of substance use and abuse among youths. The results of these studies are used to form programs and influence policies to address the substance use problem among youths.

**METHODS:** The Office of the Assistant Secretary for Planning and Evaluation was tasked with comparing and contrasting the survey design, sampling, and statistical assessment used in the three surveys: the National Household Survey on Drug Abuse (NHSDA), the Monitoring the Future (MTF) study, and the Youth Risk Behavior Survey (YRBS).

**RESULTS/CONCLUSIONS:** Although the three surveys produce different prevalence estimates for drug use and abuse among youths, all three have strong designs and have shown similar trends over time. The papers commissioned to compare the methodologies of these three studies should further aid in understanding the differences between these surveys and assist policymakers in tracking the prevalence and trends of drug use.

### **Substance use survey data collection methodologies [Commentary]**

**CITATION:** Kann, L. (2001). Substance use survey data collection methodologies [Commentary]. *Journal of Drug Issues*, 31(3), 725-728.

**PURPOSE/OVERVIEW:** Alcohol and drug use is a leading social and health concern in the United States. It is associated with and contributes to higher mortality rates and crime. Therefore, a national surveillance of alcohol and drug use is an integral public health issue.

**METHODS:** This paper is a commentary on the five papers sponsored by the Office of the Assistant Secretary for Planning and Evaluation to evaluate the differences in methodology used in the leading surveys on drug use in the United States and to assess the impact on prevalence estimates for drug use.

**RESULTS/CONCLUSIONS:** The author concluded that the five papers were a solid resource in understanding the different methodologies used in collecting information about drug use. The papers also highlighted further methodological work that could be done to clarify more fully the effects of methodological differences on the prevalence estimates for the three surveys.

### **Needs for state and local data of national relevance**

**CITATION:** Lepkowski, J. M. (2001). Needs for state and local data of national relevance. In M. L. Cynamon & R. A. Kulka (Eds.), *Seventh Conference on Health Survey Research Methods* (HHS Publication No. PHS 01-1013, pp. 247-250). Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

**PURPOSE/OVERVIEW:** Data from health surveys provide information critical to State and local governments for health policy and resource allocation decisions. The four surveys described in the session that this discussion paper addresses employ various methodological approaches for collecting health data for State and local areas.

**METHODS:** The four surveys addressed in this discussion paper include: the Behavioral Risk Factor Surveillance System (BRFSS), the National Household Survey on Drug Abuse (NHSDA),

the National Immunization Survey (NIS), and the State and Local Area Integrated Telephone Survey (SLAITS).

**RESULTS/CONCLUSIONS:** The four surveys share important design goals, such as producing valid estimates for the entire country and for States. However, the BRFSS, NIS, and SLAITS all have noncoverage concerns because of the use of telephone sampling methods for the household portion of the population. Also, although nonresponse is present in all four surveys, the potential nonresponse bias is greatest for the three surveys employing telephone sampling methods. The combination of noncoverage error and nonresponse error can be problematic, especially at the State level. The author suggests several ways to address this concern, including compensatory weights for noncoverage and nonresponse, the use of supplemental frames, and researching models to fully address factors that affect nonresponse. The author also notes that improved survey content and the coordination of State- and nationally administered surveys are of great importance to increasing the quality of State and local health surveys.

### **Development of computer-assisted interviewing procedures for the National Household Survey on Drug Abuse**

**CITATION:** Office of Applied Studies. (2001, March). *Development of computer-assisted interviewing procedures for the National Household Survey on Drug Abuse* (HHS Publication No. SMA 01-3514, Methodology Series M-3). Rockville, MD: Substance Abuse and Mental Health Services Administration.

**PURPOSE/OVERVIEW:** In 1999, the National Household Survey on Drug Abuse (NHSDA) sample was expanded and redesigned to permit using a combination of direct and model-based small area estimation (SAE) procedures that allow the Substance Abuse and Mental Health Services Administration (SAMHSA) to produce estimates for all 50 States and the District of Columbia. In addition, computer-assisted data collection procedures were adopted for both screening and interviewing. This report summarizes the research to develop these computer-assisted screening and interviewing procedures.

**METHODS:** This report covers a variety of NHSDA field experiment topics. To start, Chapter 2 gives a brief history of research on the NHSDA, and Chapter 3 offers further background information, including a literature review and an overview of critical design and operational issues. Chapter 4 focuses on the 1996 feasibility experiment and cognitive laboratory research, while Chapters 5 through 9 delve into the 1997 field experiment. Specifically, Chapter 5 summarizes the design and conduct of the 1997 effort, Chapter 6 compares computer-assisted personal interviewing (CAPI) and audio computer-assisted self-interviewing (ACASI) with paper-and-pencil interviewing (PAPI) for selected outcomes, Chapter 7 describes the effect of ACASI experimental factors on prevalence and data quality, Chapter 8 details the development and testing of an electronic screener, and Chapter 9 describes the operation of the 1997 field experiment. The next two chapters offer insights into the willingness of NHSDA respondents to be interviewed (Chapter 10) and the effect of NHSDA interviewers on data quality (Chapter 11). Chapter 12 is devoted to further refinement of the computer-assisted interviewing (CAI) procedures during the 1998 laboratory and field testing of a tobacco module.

**RESULTS/CONCLUSIONS:** The main results from the 1997 field experiment were the following: (1) Use of a single gate question to ask about substance use rather than multiple gate questions yielded higher reporting of substance use. (2) The use of consistency checks within the survey instrument yielded somewhat higher reporting of drug use than when such checks were not used. (3) The use of ACASI yielded higher reports of drug use than PAPI. (4) Respondents were less likely to request help in completing the survey in ACASI than in PAPI, particularly among youths (aged 12–17) and adults with less than a high school education. (5) Respondents using ACASI reported higher comfort levels with the interview than those using PAPI. (6) Respondents with fair or poor reading ability found the recorded voice in ACASI more beneficial than those with excellent reading ability. (7) ACASI respondents reported higher levels of privacy than PAPI respondents.

### **National Household Survey on Drug Abuse: 1999 nonresponse analysis report**

**CITATION:** Office of Applied Studies. (2001). *National Household Survey on Drug Abuse: 1999 nonresponse analysis report*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

**PURPOSE/OVERVIEW:** This report addresses the nonresponse patterns obtained in the 1999 National Household Survey on Drug Abuse (NHSDA). This report was motivated by the relatively low response rates in the 1999 NHSDA and by the apparent general trend of declining response rates in field studies. The analyses presented in this report were produced to help provide an explanation for the rates in the 1999 NHSDA and guidance for the management of future projects.

**METHODS:** The six chapters of this report provide a background for the issues surrounding nonresponse and an analysis of the 1999 NHSDA nonresponse. The first three chapters provide context with reviews of the nonresponse trends in U.S. field studies, the current academic literature, and the NHSDA data collection patterns from 1994 through 1998. Chapter 4 describes the data collection process in 1999 with a detailed discussion of design changes, summary figures and statistics, and a series of logistic regressions. Chapter 5 compares 1998 with 1999 nonresponse patterns. Chapter 6 applies the analysis in the previous chapters to a discussion of a respondent incentive program for future NHSDA work.

**RESULTS/CONCLUSIONS:** The results of this study are consistent with the conventional wisdom of the professional survey research field and the findings in survey research literature. The nonresponse can be attributed to a set of interviewer influences, respondent influences, design features, and environmental characteristics. The nonresponse followed the demographic patterns observed in other studies, with urban and high crime areas having the worst rates. Finally, efforts taken to improve the response rates were effective. Unfortunately, the tight labor market combined with the large increase in sample size caused these efforts to lag behind the data collection calendar. The authors used the results to generate several suggestions for the management of future projects. First, efforts should be taken to convert reluctant sample elements to completions. This report contains an outline for an incentive program that addresses this issue. Second, because the characteristics of field staff are among the most important correlates of nonresponse, a detailed analysis should be conducted to evaluate the most effective

designs for staffing and retention. Finally, actions should be taken to tailor the survey to regional characteristics, such as environmental and respondent characteristics, which are important predictors of response patterns.

## **Culture and item nonresponse in health surveys**

**CITATION:** Owens, L., Johnson, T. P., & O'Rourke, D. (2001). Culture and item nonresponse in health surveys. In M. L. Cynamon & R. A. Kulka (Eds.), *Seventh Conference on Health Survey Research Methods* (HHS Publication No. PHS 01-1013, pp. 69-74). Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

**PURPOSE/OVERVIEW:** Item nonresponse is one of the available indicators that can be used to identify cultural variations in survey question interpretation and response. This paper analyzes the patterns of item nonresponse across cultural groups in the United States using data from four national health surveys. The authors hypothesized that respondents from minority cultural groups would exhibit higher nonresponse to survey questions than non-Hispanic white respondents.

**METHODS:** For the analysis, the authors used four national health-related surveys: the 1992 Behavioral Risk Factor Surveillance System (BRFSS), the 1992 National Household Survey on Drug Abuse (NHSDA), the 1991 National Health Interview Drug Use Supplement (NHIS), and the 1990–1991 National Comorbidity Survey (NCS). In each dataset, the authors selected several items that reflected different health domains. From these items, the authors created 10 dichotomous variables that measured whether the source items contained missing data. Each measure was examined using simple cross-tabulations and logistic regression models in which the authors controlled for sociodemographic variables associated with item nonresponse (e.g., age, gender, education, and marital status).

**RESULTS/CONCLUSIONS:** Although item nonresponse rates of health survey questions appeared to be low, the authors found that item nonresponse may vary systematically across cultural groups. The authors also found higher item nonresponse rates among each of the minority racial and ethnic groups as compared with non-Hispanic white respondents. In addition to these general findings, the authors noted group-specific cultural differences that warrant special attention. First, the largest odds ratio associated with respondent culture, which reflected Hispanic refusals to answer questions related to their social relationships, may be a consequence of a cultural difference: showing unwillingness to report anything other than positive relations with their family and friends. Second, greater reluctance to report substance use by African-American respondents may be explained by this group's beliefs about selective enforcement of drugs laws against minority groups in the United States. Finally, although the generally low prevalence of item nonresponse in the data analyzed may indicate that differential rates of nonresponse are of insignificant magnitude to seriously bias survey findings, the data do suggest that cultural differences in item nonresponse may be more problematic under certain conditions.

## **Person-pair sampling weight calibration using the generalized exponential model for the National Household Survey on Drug Abuse**

CITATION: Penne, M. A., Chen, P., & Singh, A. C. (2001). Person-pair sampling weight calibration using the generalized exponential model for the National Household Survey on Drug Abuse. In *Proceedings of the 2001 Joint Statistical Meetings, American Statistical Association, Social Statistics Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: For pair data analysis, sampling weights need to be calibrated such that the problems of extreme weights are addressed. Extreme weights, which give rise to high unequal weighting effect (UWE) and hence instability in estimation, are derived from possible small pairwise selection probabilities and domain-specific multiplicity factors.

METHODS: The recently developed generalized exponential model (GEM) for weight calibration at RTI by Folsom and Singh (2000) allows for multiple calibration controls, as well as separate bounds on the weight adjustment factors for extreme weights identified before calibration. Thus, controls corresponding to the number of pairs in various demographic groups from the first phase of screened dwelling units, and controls for the number of individuals in the domains of interest from the second phase of surveyed pairs and single individuals, for a key set of domains, can all be incorporated simultaneously in calibration, giving rise to a final single set of calibration weights. Numerical examples of GEM calibration of pair data and the resulting estimates for the 1999 National Household Survey on Drug Abuse (NHSDA) are presented.

RESULTS/CONCLUSIONS: Both model groups of region pair samples of the 1999 NHSDA are used to illustrate results obtained by utilizing the GEM methodology to calibrate a final analytic weight. The authors present summary results of before and after each pair adjustment step in the calibration process for the Northeast and South regions and the Midwest and West regions, respectively. Note that "after" results of each preceding adjustment step is synonymous with "before" results of every subsequent step. Sample sizes; the UWE; unweighted, weighted, and outwinsor percentages of extreme values; and the quartile distribution of both the weight component itself and the weight product up through that step also are presented.

## **How do response problems affect survey measurement of trends in drug use?**

CITATION: Pepper, J. V. (2001). How do response problems affect survey measurement of trends in drug use? In S. K. Goldsmith, T. C. Pellmar, A. M. Kleinman, W. E. Bunney, & Commission on Behavioral and Social Sciences and Education (Eds.), *Informing America's policy on illegal drugs: What we don't know keeps hurting us* (pp. 321-348). Washington, DC: National Academy Press.

PURPOSE/OVERVIEW: Two databases are widely used to monitor the prevalence of drug use in the United States. The Monitoring the Future (MTF) study surveys high school students, and the National Household Survey on Drug Abuse (NHSDA) surveys the noninstitutionalized residential population aged 12 or older. Each year, respondents from these surveys are drawn from known populations—students and noninstitutionalized people—according to well-specified

probabilistic sampling schemes. Hence, in principle, these data can be used to draw statistical inferences on the fractions of the surveyed populations who use drugs. It is inevitable, however, for questions to be raised about the quality of self-reports of drug use. Two well-known response problems hinder one's ability to monitor levels and trends: nonresponse, which occurs when some members of the surveyed population do not respond, and inaccurate response, which occurs when some surveyed individuals give incorrect responses to the questions posed. These response problems occur to some degree in almost all surveys. In surveys of illicit activity, however, there is more reason to be concerned that decisions to respond truthfully, if at all, are motivated by respondents' reluctance to report that they engage in illegal and socially unacceptable behavior. To the extent that nonresponse and inaccurate response are systematic, surveys may yield invalid inferences about illicit drug use in the United States.

**METHODS:** To illustrate the inferential problems that arise from nonresponse and inaccurate response, the author suggested using the MTF and the NHSDA to draw inferences on the annual prevalence of use for adolescents. Annual prevalence measures indicate use of marijuana, cocaine, inhalants, hallucinogens, heroin, or nonmedical use of psychotherapeutics at least once during the year. Different conclusions about levels and trends might be drawn for other outcome indicators and for other subpopulations.

**RESULTS/CONCLUSIONS:** The author concluded that the MTF and the NHSDA provide important data for tracking the numbers and characteristics of illegal drug users in the United States. Response problems, however, continued to hinder credible inference. Although nonresponse may have been problematic, the lack of detailed information on the accuracy of response in the two national drug use surveys was especially troubling. Data were not available on the extent of inaccurate reporting or on how inaccurate response changes over time. In the absence of good information on inaccurate reporting over time, inferences on the levels and trends in the fraction of users over time were largely speculative. It might be, as many had suggested, that misreporting rates were stable over time. It also might have been that these rates varied widely from one period to the next. The author indicated that these problems, however, did not imply that the data were uninformative or that the surveys should be discontinued. Rather, researchers using these data must either tolerate a certain degree of ambiguity or must be willing to impose strong assumptions. The author suggested practical solutions to this quandary: If stronger assumptions are not imposed, the way to resolve an indeterminate finding is to collect richer data. Data on the nature of the nonresponse problem (e.g., the prevalence estimate of nonrespondents) and on the nature and extent of inaccurate response in the national surveys might be used to both supplement the existing data and to impose more credible assumptions. Efforts to increase the valid response rate may reduce the potential effects of these problems.

### **Predictive mean neighborhood imputation with application to the person-pair data of the National Household Survey on Drug Abuse**

**CITATION:** Singh, A., Grau, E., & Folsom, R., Jr. (2001). Predictive mean neighborhood imputation with application to the person-pair data of the National Household Survey on Drug Abuse. In *Proceedings of the 2001 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** The authors present a simple method of imputation termed "predictive mean neighborhood," or PMN. Features of this method include the following: (1) it allows for several covariates; (2) the relative importance of covariates is determined objectively based on their relationship to the response variable; (3) it incorporates design weights; (4) it can be multivariate in that correlations between several imputed variables are preserved, as well as correlations across imputed and observed variables; (5) it accommodates both discrete and continuous variables for multivariate imputation; and finally (6) it should lend itself to a simple variance estimation adjusted for imputation.

**METHODS:** The PMN method is a combination of prediction modeling with a random nearest neighbor hot deck. It uses a model-based predictive mean to find a small neighborhood of donors from which a single donor is selected at random. Thus, the residual distribution is estimated nonparametrically from the donors, where the imputed value is (approximately) the predictive mean plus a random residual. Applications of PMN for imputing two types of multiplicity factors required for pair data analysis from the 1999 NHSDA are discussed.

**RESULTS/CONCLUSIONS:** The PMN methodology has been widely used for the imputation of a variety of variables in the NHSDA, including both continuous and categorical variables with one or more levels. The models were fit using standard modeling procedures in SAS<sup>®</sup> and SUDAAN<sup>®</sup>, while SAS macros were used to implement the hot-deck step, including the restrictions on the neighborhoods. Although creating a different neighborhood for each item nonrespondent was computationally intensive, the method was implemented successfully. At the time this paper was presented, the imputations team at RTI was implementing a series of simulations to evaluate the new method, comparing it against the unweighted sequential hot deck used earlier and a simpler model-based method.

## **Examining substance abuse data collection methodologies**

**CITATION:** Sudman, S. (2001). Examining substance abuse data collection methodologies. *Journal of Drug Issues*, 31(3), 695-716.

**PURPOSE/OVERVIEW:** The U.S. Department of Health and Human Services (HHS) sponsors three national surveys on drug use in adolescents. These three studies are the National Household Survey on Drug Abuse (NHSDA), the Youth Risk Behavior Surveillance System (YRBSS), and the Monitoring the Future (MTF) study. The estimates for drug use reported for three surveys differ considerably. This paper examines methodological reasons for the differences reported in these studies.

**METHODS:** The author does not conduct any new experiments or analyses on the survey data, but examines previous validation and methodological research. The major differences in methodology identified, which might contribute to differences in estimates, are mode of administration (home vs. school setting), questionnaire context and wording, sample design, and weighting.

**RESULTS/CONCLUSIONS:** Of all the methodological differences, it appears that the context and introduction to the questionnaires in addition to the mode of administration have the largest

impact because they affect the respondents' perceived anonymity. It is likely that NHSDA respondents do not perceive the study to be as anonymous as respondents in both the MTF and YRBSS. More research should be conducted to ascertain how anonymous respondents felt the study was.

## 2002

### **Assessment of the computer-assisted instrument**

CITATION: Caspar, R., & Penne, M. (2002). Assessment of the computer-assisted instrument. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 53-84). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

PURPOSE/OVERVIEW: The conversion of the National Household Survey on Drug Abuse (NHSDA) to computer-assisted interviewing (CAI) offered an opportunity to improve the quality of the data collected in the NHSDA in a number of ways. Some of these improvements were implemented easily and manifested themselves in more complete data (e.g., the ability to eliminate situations where questions were inadvertently left blank by the respondent). However, other improvements could only be realized through careful development and implementation of new procedures. Thorough testing was needed to determine whether these new procedures did, in fact, result in higher quality data.

METHODS: This chapter describes two significant revisions to how key NHSDA data items are collected and the effect of these revisions on the quality of the data obtained in the 1999 NHSDA. The first of these revisions was the addition of a methodology for resolving inconsistent or unusual answers provided by the respondent. This methodology was incorporated into the collection of a large number of the data items that are considered critical to the reporting needs of the NHSDA. The second revision dealt specifically with the way data on frequency of substance use over the past 12-month period was reported. This chapter also provides a review of several basic measures of data quality, including rates of "Don't Know" and "Refused" responses, breakoff interviews, and the observational data provided by the interviewers at the conclusion of each interview. Where possible, these measures are compared between the CAI and paper-and-pencil interviewing (PAPI) NHSDA instruments as a means of assessing the effect of the move to CAI on data quality.

RESULTS/CONCLUSIONS: The results suggested that the move to CAI data collection has improved data quality, although in some cases the increase was fairly small because data quality for the PAPI NHSDA data was already quite high. Perhaps the most significant improvement to data quality came as a result of the inclusion of the inconsistent and unusual data checks (described in Section 4.1). This enhancement was a radical departure from the PAPI NHSDA and one that was possible only under the CAI mode of interview. The results presented here show that, although the CAI data did not suffer from a large amount of inconsistent or unusual data, the methodology was able to resolve a large number of these cases in a way that was both cost-effective and likely to enhance overall data quality for the items involved. Results of the change in the 12-month frequency of use item were somewhat difficult to interpret. The revised method for collecting these data resulted in higher reported frequencies, but whether this was due to the revision or simply the move from PAPI to CAI was impossible to determine. The distribution of responses and the fact that the mean frequency was higher under CAI than PAPI provided anecdotal support for the revised method of collecting these data, however. Finally,

results from basic measures of data quality and interviewer debriefing items suggested that the CAI methodology reduced interview difficulties among respondents, helped to further enhance the degree of privacy, and appeared to contribute positively to item-level response rates.

## **Mode effects on substance use measures: Comparison of 1999 CAI and PAPI data**

**CITATION:** Chromy, J., Davis, T., Packer, L., & Gfroerer, J. (2002). Mode effects on substance use measures: Comparison of 1999 CAI and PAPI data. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 135-159). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** The shift from paper-and-pencil interviewing (PAPI) to computer-assisted interviewing (CAI) in the 1999 National Household Survey on Drug Abuse (NHSDA) was anticipated to have significant effects on the reporting of substance use by survey respondents. This expectation was based on several studies in the literature, as well as field testing done within the NHSDA project that showed respondents were more willing to report sensitive behaviors using audio computer-assisted interviewing than with self-administered paper questionnaires. Because of the great interest in analyzing trends in substance use prevalence, a critical component of the 1999 NHSDA redesign was the supplemental sample that employed the "old" PAPI NHSDA data collection methodology. The intent of this dual-sample design in 1999 was primarily to make it possible to continue to measure trends, using estimates from both before and after the redesign. However, the large dual-sample design of the 1999 NHSDA also can be viewed as an important survey research experiment assessing mode effects on the reporting of sensitive behaviors. With its dual-sample design, the 1999 NHSDA provided a large sample for assessing the impact of mode of interview. However, the intention of the 1999 NHSDA design was not to evaluate the mode effect, but to determine the impact of the overall change in method due to the redesign of the survey in 1999 on the time series of substance use statistics. The overall change involved many aspects of the survey design and estimation procedures in addition to mode, such as the sampling plan, the questionnaire, data editing, and imputation. Isolating the "pure" mode effect with these data was difficult and was complicated further by the unexpected impact of interviewer experience on substance use prevalence estimates (see Chapter 8 of the 2002 report). Nevertheless, the analyses presented in this chapter provide some important findings concerning mode effects, as well as on the comparability of pre-1999 NHSDA published estimates with estimates from the redesigned NHSDA. When studying the reporting of sensitive or illegal behaviors, conventional thinking has been that higher reporting is closer to the truth. This evaluation continued that approach, recognizing that this may not be true in every case.

**METHODS:** The authors compare the substance use prevalence estimates derived from the 1999 PAPI and CAI samples. In addition to providing NHSDA data users with information that will help them interpret NHSDA trends in substance use prevalence, the analysis also is of interest to survey researchers concerned with mode effects.

**RESULTS/CONCLUSIONS:** The results support previous research that shows higher reporting of sensitive behaviors with audio computer-assisted self-interviewing (ACASI) than with self-administered paper answer sheets. A total of 336 comparisons of unedited estimates from PAPI and CAI were made. Of these, 112 indicated significantly higher CAI estimates, while only 5 indicated significantly higher PAPI estimates. Higher CAI estimates were particularly evident for lifetime prevalence, the measures impacted least by questionnaire structure differences between CAI and PAPI. The analyses of edited and imputed NHSDA estimates showed mixed results for past year and past month use. Lifetime prevalence rates, which were minimally affected by editing and imputation, were generally higher with CAI than with PAPI. A total of 448 comparisons of edited and imputed estimates from PAPI and CAI were made. Of a total of 62 statistically significant differences in lifetime prevalence, 56 indicated a higher CAI estimate and 6 indicated a higher PAPI estimate. Results for past year and past month measures showed variation across substances and age groups and also were different in the full sample analysis than in the matched sample analysis. Out of 87 significant differences for past month or past year use, 41 indicated a higher CAI estimate and 46 indicated a higher PAPI estimate. For past month use of alcohol and cigarettes, the substances with the highest prevalence of use, the observed PAPI estimate was higher than the CAI estimate for every age group and for both the full and matched sample analyses, and 11 of the 16 comparisons were statistically significant. Marijuana use estimates also tended to be higher with PAPI, although most of these differences were not statistically significant. One clear conclusion is that the CAI mode of interviewing led to more internally consistent and complete data. Under PAPI, the need for editing and imputation to clarify recency of use was larger for most substances and provided a greater opportunity to influence estimates through the editing and imputation process. In summary, the CAI methodology produced more complete data with a lower requirement for editing and higher prevalence estimates when treating only unambiguous reports as positive indications of substance use.

### **Nonresponse in the 1999 NHSDA**

**CITATION:** Eyerman, J., Odom, D., Wu, S., & Butler, D. (2002). Nonresponse in the 1999 NHSDA. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 23-51). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** The redesign of the National Household Survey on Drug Abuse (NHSDA) in 1999 resulted in major changes in many aspects of the data collection procedures. This raised concerns that the response rates could be affected. In particular, the increased sample size, reduced clustering of sample segments, transition from a paper to a computerized screening instrument, and the transition from paper-and-pencil interviewing (PAPI) to computer-assisted interviewing (CAI) all had the potential to change the response rates. During the first quarter of 1999, an assessment of the progress in completing the fieldwork indicated a reduction in the response rates, relative to response rates achieved historically in the NHSDA. To address this problem, several management actions were implemented immediately. Although the response rates improved steadily throughout the remainder of the year, the result was significantly lower response rates for the 1999 NHSDA than for prior NHSDAs.

**METHODS:** Extensive analysis was undertaken in an attempt to understand the reasons for the drop in response rates and how it was related to each of the design changes. This chapter summarizes this analysis. It also discusses the management actions implemented during 1999 to improve the response rates and assesses the effectiveness of these actions.

**RESULTS/CONCLUSIONS:** The design changes between 1998 and 1999 corresponded with a large decrease in the response rates. A series of management efforts was taken to address the decrease, both in anticipation of the design changes and in reaction to unexpected results of the changes. In general, the efforts were successful. The extended analysis presented in this chapter summarizes the current understanding of the decline in the response rates in 1999. First, it appears that the previous understanding of the correlates of nonresponse was correct, but it does not completely explain the difference between 1998 and 1999. Second, management efforts taken during 1999 appear to have been successful in reducing the decline in the response rates, and this success carried over to 2000. Third, the computerized screening instrument reduced sampling bias by removing interviewer effects from the screening routine, and this had a small and negative impact on the response rates. Fourth, the transition from PAPI to CAI increased the response rates. Finally, much of the decline in 1999 can be attributed to changes in the composition of the field staff resulting from the large increase in sample size, most notably the reduced number of experienced field interviewers working on the project. However, this does not fully explain the decline in 1999.

### **Introduction. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues***

**CITATION:** Gfroerer, J. (2002). Introduction. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 1-8). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** In 1999, a major redesign of the National Household Survey on Drug Abuse (NHSDA) was implemented involving both the sample design and the data collection method of the survey. The strictly national design was changed to a much larger, State-based design to meet the needs of policymakers for estimates of substance use prevalence for each State. The data collection method was changed from a paper-and-pencil interviewing (PAPI) method to a computer-assisted interviewing (CAI) method, primarily to improve the quality of NHSDA estimates. This report has two purposes. First, it provides information on the impact of the redesign on the estimates produced from NHSDA. Researchers and other users of NHSDA data will find this information helpful in interpreting NHSDA estimates, particularly if they are interested in comparing data from the new design with data from the old design. The second purpose is to present research findings of interest to survey methodologists involved in designing and conducting surveys of all types, not just surveys of substance abuse.

**METHODS:** Implementation of these significant changes posed a number of difficult challenges involving a variety of methodological issues. These issues cover many aspects of the survey, including the management of fieldwork, the processing of data, and the reporting of results.

This publication discusses several of the most critical issues encountered and describes how the research team conducting the survey addressed them.

**RESULTS/CONCLUSIONS:** Although these findings taken as a whole could be considered a case study in the redesign of a major ongoing survey, several of the chapters in this report present important research findings that are applicable to many types of surveys.

### **Redesigning an ongoing national household survey: Methodological issues**

**CITATION:** Gfroerer, J., Eyerman, J., & Chromy, J. (Eds.). (2002). *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** In 1999, a major redesign of the National Household Survey on Drug Abuse (NHSDA) was implemented involving both the sample design and the data collection method of the survey. The data collection method was changed from a paper-and-pencil interviewing (PAPI) method to a computer-assisted interviewing (CAI) method, primarily to improve the quality of NHSDA estimates. Implementation of these significant changes posed a number of difficult challenges involving a variety of methodological issues. These issues cover many aspects of the survey, including the management of fieldwork, the processing of data, and the reporting of results.

**METHODS:** This publication discusses several of the most critical issues encountered and describes how the research team conducting the survey addressed them. This report has two purposes. First, it provides information on the impact of the redesign on the estimates produced from NHSDA. Researchers and other users of NHSDA data will find this information helpful in interpreting NHSDA estimates, particularly if they are interested in comparing data from the new design with data from the old design. The second purpose is to present research findings of interest to survey methodologists involved in designing and conducting surveys of all types, not just surveys of substance abuse. Although these findings taken as a whole could be considered a case study in the redesign of a major ongoing survey, several of the chapters in this report present important research findings that are applicable to many types of surveys.

**RESULTS/CONCLUSIONS:** N/A.

### **Impact of interviewer experience on respondent reports of substance use**

**CITATION:** Hughes, A., Chromy, J., Giacoletti, K., & Odom, D. (2002). Impact of interviewer experience on respondent reports of substance use. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 161-184). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** The redesign of the National Household Survey on Drug Abuse (NHSDA) in 1999 included a change in the primary mode of data collection from paper-and-

pencil interviewing (PAPI) to computer-assisted interviewing (CAI). In addition, the sample design changed from a Nation-based design to a State-based one, and a supplemental sample was collected using PAPI to measure change between 1999 and earlier years. The overall sample size increased from 25,500 in 1998 to 80,515 in 1999, including the CAI and supplemental PAPI samples. Consequently, it was necessary to hire more interviewers than in previous years, which resulted in a higher proportion of inexperienced interviewers. New interviewing staff turnover also was high in 1999, requiring additional training of newly hired interviewers and contributing to the general inexperience of the interviewing staff for both the CAI and PAPI samples.

**METHODS:** This chapter describes the analysis that was done to understand and explain the relationship between the effects of changes in mode, sample design, and interviewer staffing.

**RESULTS/CONCLUSIONS:** The analysis presented in this chapter indicates that the uneven mix of experienced and inexperienced NHSDA field interviewers (FIs) in 1999 had some effect on estimated substance use rates for that year. Overall, the effect on 1999 CAI prevalence estimates was smaller in magnitude than the effect on 1999 PAPI rates, which was an indication that the CAI methods played a role in reducing the effects of FI experience on substance use rates. However, because the mechanism of these effects was unknown, it was determined that additional studies would be undertaken to increase an understanding of this phenomenon. In the meantime, analyses of interviewer effect as seen in this chapter were to continue to be presented in subsequent reports. These findings resulted in an added emphasis—in training and in the field—on encouraging experienced and new FIs to follow the interview protocol.

## **Development of editing rules for CAI substance use data**

**CITATION:** Kroutil, L., & Myers, L. (2002). Development of editing rules for CAI substance use data. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 85-109). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** A major change to the study protocol for the 1999 National Household Survey on Drug Abuse (NHSDA) was the shift from paper-and-pencil interviewing (PAPI) to computer-assisted interviewing (CAI). Although many of the substance use questions are similar in the two instruments, there are some differences. In addition, whereas the PAPI questionnaire required respondents to answer all questions in most sections, the CAI instrument makes extensive use of skip instructions. These significant differences in the nature of the data obtained in the new and old instruments necessitated the development of entirely new editing rules.

**METHODS:** This chapter discusses the development of the new editing rules for the NHSDA CAI data and presents the results of an investigation of alternative editing methods for CAI data. The analysis was based primarily on data from the first 6 months of data collection in 1999. The authors discuss data quality issues that were affected by conversion to CAI, present the general methodological approach used to define and test alternative rules for defining a usable case, discuss alternative editing rules, along with the final rule that was implemented, and

present selected substance use measures to compare the impact of the new CAI editing procedures with the former PAPI procedures.

**RESULTS/CONCLUSIONS:** The authors find for some substances that there was little change in going from the raw CAI data provided by the respondents to the final imputed estimates. For example, the estimate of marijuana use in the past month changed from 4.6 percent in the raw data to 4.7 percent after editing and imputation. In comparison, editing made a greater contribution to estimates of past month marijuana use and cigarette use in the 1998 data. For past month marijuana use, the raw estimate in 1998 was 4.0 percent (weighted), the estimate after editing was 5.0 percent, and the final imputed estimate was 5.0 percent. Thus, the editing procedures that had been used in the NHSDA since 1994 increased the 1998 estimate of past month marijuana use by about 25 percent relative to the raw data; the additional impact of imputation on the final estimate of past month marijuana use was virtually nil. Differences in the impacts of the 1999 CAI editing and imputation procedures and those used in prior years were even more pronounced for less commonly used substances. Following imputation, a total of 70 CAI respondents were classified as past month heroin users, or a net increase of only 4 cases relative to the raw and a net increase of 6 relative to the edited. In comparison, the editing procedures in 1998 nearly doubled the number of respondents classified as being past month heroin users (17 respondents in the raw data and an additional 11 cases who were assigned to this category through editing). In all, these changes in the 1999 CAI editing procedures represent an improvement over the way missing or inconsistent data had been handled in that these issues are resolved primarily through statistical methods.

### **2001 National Household Survey on Drug Abuse: Incentive experiment combined quarter 1 and quarter 2 analysis**

**CITATION:** Office of Applied Studies. (2002, July). *2001 National Household Survey on Drug Abuse: Incentive experiment combined quarter 1 and quarter 2 analysis* (RTI 07190.388.100, prepared for the Office of Applied Studies, Substance Abuse and Mental Health Services Administration, by RTI under Contract No. 283-98-9008). Rockville, MD: Substance Abuse and Mental Health Services Administration.

**PURPOSE/OVERVIEW:** The purpose of this report was to summarize the results of the incentive experiment in the 2001 National Household Survey on Drug Abuse (NHSDA) and to evaluate the best treatment option for the use of monetary incentives in future NHSDAs. The NHSDA experienced a considerable decline in response rates in 1999 due in part to the transition from a national probability sample to a State probability sample designed to yield State-level estimates. A series of management adjustments were made to improve the response rates in 2000. In general, the adjustments were successful, and a recovery was made from the 1999 decline. However, the rates remained below the project target rate and the historical NHSDA average. An incentive given to respondents was considered as an option for addressing the downward trend in respondent cooperation. However, it has been noted that incentives may have a negative impact on areas of data quality other than unit response rates (Shettle & Mooney, 1999). Although it may lead to better response rates, it is possible that the additional costs may exceed the constraints of the project budget. In an effort to understand the risks and benefits associated with a respondent incentive, the NHSDA's sponsor, the Substance Abuse and Mental

Health Services Administration (SAMHSA), requested that the NHSDA's contractor, Research Triangle Institute (RTI), conduct a special methodological field test in the form of an incentive experiment.

**METHODS:** The experiment was overlaid on the NHSDA main study data collection sample and scheduled during the first two quarters of 2001. A randomized, split-sample, experimental design was included with the main study data collection of the NHSDA to compare the impact of \$20 and \$40 incentive treatments with a \$0 control group on measures of respondent cooperation, data quality, survey costs, and population substance use estimates. This report is the second of two. The first report describes the experimental design and the results from data collection in the first quarter of 2001 (Eyerman, Bowman, Odom, Vatalaro, & Chromy, 2001a). This second report provides combined findings for the full experiment for both quarters.

**RESULTS/CONCLUSIONS:** The results were very promising. The \$20 and the \$40 treatments produced significantly better interview response rates than the control for the combined results of both quarters of the experiment. This improvement led to a gain in overall response rates of about 10 points for each treatment. Furthermore, both the \$20 and the \$40 treatments more than paid for themselves, each resulting in a lower data collection cost per completed case, including the incentive, than the control. The incentives had a favorable impact on measures of respondent cooperation. Both treatments had significantly lower refusal rates than the control's rate, and the \$40 treatment had significantly lower noncontact rates than the control's. Field interviewers reported that the incentives reduced the amount of effort required to complete a case and that the incentives influenced the respondent's decision to cooperate. Perhaps most importantly, the incentives had little impact on the population estimates of past month alcohol, cigarette, or marijuana use. The prevalence rates for past month use of these substances by respondents in the treatment groups were not significantly different from those reported by those in the control. This suggests that incentives encourage greater participation by respondents, but do not change their self-reported substance use. Incentives may thus improve estimates by reducing nonresponse bias without increasing response bias. Taken together, the results clearly favor a \$40 incentive for all individuals selected for the NHSDA.

### **Changes in NHSDA measures of substance use initiation**

**CITATION:** Packer, L., Odom, D., Chromy, J., Davis, T., & Gfroerer, J. (2002). Changes in NHSDA measures of substance use initiation. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 185-220). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** The National Household Survey on Drug Abuse (NHSDA) data are used to generate a number of annual estimates relating to the initiation of substance use. Using the responses to retrospective questions about age of first use, annual estimates are generated for the incidence rate of first substance use, for the number of initiates to substance use, and for the average age of first use. Estimates of new initiates and average age of first use are reported for all lifetime substance users aged 12 or older. These use initiation measures are important because they can capture the rapidity with which new substance users arise in specific population

subgroups and can identify emerging patterns of substance use. The redesign of the NHSDA in 1999 introduced some changes in the questions about initiation, as well as the method of administration. In the presence of these changes, the overall data processing and estimation methodologies were reviewed and, in some cases, revised. (The revisions to the editing and imputation procedures are summarized in Chapters 5 and 6 of this 2002 report.) Limitations of the existing methodology for computing incidence rates were found. As a result, a new incidence rate methodology was developed. The definition of initiation of daily cigarette use was modified, and an adjustment to the program logic in the calculation for the incidence of first daily use of cigarettes also was made.

**METHODS:** This chapter is organized in three sections addressing the impact of methodological change on substance use initiation measures. Section 9.1 describes the old and new incidence rate estimation methods and evaluates its impact; the impact of the editing and imputing changes is evaluated in conjunction with the method impact. Section 9.2 focuses on the questionnaire wording and administration mode effects. Section 9.3 focuses on all the issues associated with initiation of first daily use of cigarettes.

**RESULTS/CONCLUSIONS:** Although the estimates for individual years were quite variable, the overall average impact of the new editing and imputation procedures was to increase incidence rates for both age groups (12 to 17 and 18 to 25) and to increase the estimated number of new initiates. The largest impacts were observed for pain relievers and other substances that use multiple gate questions before presenting the age-of-first-use question. Estimates of the average age of initiation did not appear to be consistently changed in either direction by the change in editing and imputation. The impact of the new method of incidence rate calculation also was studied. The number of new initiates occurring at age 17 was presumably quite high for almost all substances. The new incidence rate calculation rules treated respondents as 17 year olds right up to (but not including) their 18th birthday. The old rule classified respondents as 18 years old for the entire year in which their 18th birthday occurred. Thus, the new calculation method had the effect of increasing the estimates of time at risk and the number of initiates for 17 year olds, but because the number of initiates is high at age 17, the overall impact was greater on the numerator than the denominator. As a result, the incidence rates for youths aged 12 to 17 increased and the incidence rates for adults 18 to 25 usually decreased somewhat with the new method. Mode effects could not be cleanly isolated because of some accompanying changes in the question routing process and supplementary questions on date of first use for recent users that were implemented in conjunction with the implementation of computer-assisted interviewing (CAI). Within this limitation, comparable data from paper-and-pencil interviewing (PAPI) and CAI were studied. One somewhat surprising result was that the level of missing or inconsistent data actually increased with the introduction of CAI. However, this may have resulted because of the increased number of checks employed to identify inconsistent data in the post-survey processing. The increase in the proportion of missing age-at-first-use data may have been facilitated by the respondent's option to answer "Don't know" or "Refused" in CAI. A pattern of mode effects similar to that observed for reported lifetime substance use was found, with generally higher reporting of initiation in CAI than in PAPI.

## **Predictive mean neighborhood imputation for NHSDA substance use data**

**CITATION:** Singh, A., Grau, E., & Folsom, R., Jr. (2002). Predictive mean neighborhood imputation for NHSDA substance use data. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 111-133). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** In 1999, the instrument used to administer the National Household Survey on Drug Abuse (NHSDA) was changed from a paper-and-pencil interviewing (PAPI) format to a computer-assisted interviewing (CAI) format. In previous years, imputation of missing values in recency of use and frequency of use in the past 12 months was accomplished with an unweighted sequential hot-deck procedure. In the spirit of improving the quality of estimates from the redesigned NHSDA and as a result of fundamental differences between PAPI and CAI, there was a need to change the way missing data were edited and imputed. The implementation of the "flag and impute" editing rule, described in this 2002 report's Chapter 5, and the desire to impute more variables required a new method that was rigorous, flexible, and preferably multivariate.

**METHODS:** This chapter presents a new imputation method with these characteristics, termed predictive mean neighborhoods (PMN), that was used to impute missing values in the NHSDA substance use variables. Following a discussion of background in Section 6.1, this chapter outlines the previously used hot-deck method, along with its limitations, in Section 6.2. The new method is described in general in Section 6.3, followed by details of the method in Section 6.4. Section 6.5 compares the method with other available methods and provides details concerning the motivation for employing a new method. In the concluding section (Section 6.6), the impact of imputation on substance use estimates is compared between PAPI and CAI.

**RESULTS/CONCLUSIONS:** The authors assess the relative impact of imputation for 1998 and 1999 (CAI) estimates of past month use, past year use, and lifetime use of all substances in the core section of the questionnaire. Because of numerous changes between the 1998 sample and the 1999 CAI sample, it would not be advisable to compare the final prevalence estimates (final percent) between the two samples. However, some comments can be made about the comparison of the "relative percent from imputes" between the two samples. In general, imputation had greater impact on the prevalence estimates in the CAI sample than on the estimates in the PAPI sample. With the implementation of the flag-and-impute editing rule in the CAI sample, where inconsistencies would be resolved by imputation, this result was not surprising. The exceptions to this rule were either due to differences in questionnaire format between PAPI and CAI or to attributes of the modules themselves.

## **1999–2001 National Household Survey on Drug Abuse: Changes in race and ethnicity questions**

CITATION: Snodgrass, J. A., Grau, E. A., & Caspar, R. A. (2002, October 18). *1999–2001 National Household Survey on Drug Abuse: Changes in race and ethnicity questions* (RTI 07190.488.010, prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, Contract No. 283-98-9008). Research Triangle Park, NC: RTI.

PURPOSE/OVERVIEW: Since the inception of the National Household Survey on Drug Abuse (NHSDA), renamed the National Survey on Drug Use and Health (NSDUH) as of 2002, questions have been included to determine the race and ethnicity of each respondent. Race and ethnicity are routinely used as part of the demographic breakdowns in the analyses and the various reports generated from the survey. From 1971 to 1998, the race and ethnicity questions underwent few changes. (See Appendix A of this 2002 report for the full list of race and ethnicity questions used for each NHSDA survey year from 1971 to 2001.) However, along with the switch from paper-and-pencil interviewing (PAPI) methods of questionnaire administration to computer-assisted interviewing (CAI) methods in 1999, the race and ethnicity categories were updated pursuant to new Office of Management and Budget (OMB) directives.

METHODS: This report details the revisions to the race and ethnicity questions. The report includes the history of the change, how the change affected the editing and imputation procedures, and how it changed the derivation of the race and ethnicity variables used in NHSDA analyses.

RESULTS/CONCLUSIONS: N/A.

## **Substance abuse among older adults in 2020: Projections using the life table approach and the National Household Survey on Drug Abuse**

CITATION: Woodward, A. (2002, December). Substance abuse among older adults in 2020: Projections using the life table approach and the National Household Survey on Drug Abuse. In S. P. Korper & C. L. Council (Eds.), *Substance use by older adults: Estimates of future impact on the treatment system* (HHS Publication No. SMA 03-3763, Analytic Series A-21, pp. 95-105). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

PURPOSE/OVERVIEW: One way of projecting substance use problems among older adults is to use a life table approach. The National Household Survey on Drug Abuse (NHSDA), a major data source on substance use and abuse among the U.S. civilian population aged 12 or older, could potentially be used in a life table approach.

METHODS: In the life table approach, cohorts are followed for a given period of time to determine their various outcomes. In the table, one cohort's drug use is followed as the cohort ages. The NHSDA data are reviewed to see whether they can be used in this fashion to produce estimates for the groups who used illicit drugs or drank heavily or who were substance dependent.

**RESULTS/CONCLUSION:** A review of the NHSDA shows that, even with its increase in sample size in 1999, the survey currently does not provide sufficient detailed data to be used in a life table approach. The survey could be expanded, however, with selected questions added in a special supplement so that a life table or other more sophisticated approach could be used to make projections of substance use problems among older adults.

### **Summary of NHSDA design changes in 1999**

**CITATION:** Wright, D., Barker, P., Gfroerer, J., & Piper, L. (2002). Summary of NHSDA design changes in 1999. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 9-22). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** An entirely new sample design and a state-of-the-art data collection methodology were implemented with the 1999 National Household Survey on Drug Abuse (NHSDA). The sample design changed from a national, stratified, multistage area probability sample to a 50-State design, with independent stratified, multistage area probability samples selected in each State. The sample size increased from about 25,500 interviews in 1998 to about 67,000 interviews in 1999. For the first time in NHSDA history, the 1999 survey administered the interview using computer-assisted interviewing (CAI) technology exclusively, including both computer-assisted personal interviewing (CAPI) and audio computer-assisted self-interviewing (ACASI). Because this new methodology was being implemented, an additional national sample was selected, and about 14,000 interviews administered using the previous paper-and-pencil interviewing (PAPI) methodology.

**METHODS:** Together, the PAPI sample and the CAI sample served three purposes. First, the PAPI samples for 1998 and 1999 provided a way to continue to measure the trend in substance use for that period. Second, with both representative samples for 1999, the effect of the change in data collection from PAPI to CAI could be measured without being confounded with the measurement of trends. Third, with a measurement of the impact of the switch to the CAI methodology, estimates for 1998 and earlier years could be adjusted to be comparable with CAI estimates for 1999 and later so that long-term trends in substance use could be estimated.

The CAI and PAPI samples for 1999 together resulted in 81,000 completed interviews. The 1999 NHSDA fully employed another technological innovation: use of a hand-held computer at each sample dwelling unit to conduct household screening and to select the sample person(s) for the interview. With this new design, technology, and markedly increased sample size, the structure of the data collection staff also had to be modified significantly for 1999. This chapter presents details of these changes.

**RESULTS/CONCLUSIONS:** N/A.

## 2003

### **Possible age-associated bias in reporting of clinical features of drug dependence: Epidemiological evidence on adolescent-onset marijuana use**

CITATION: Chen, C. Y., & Anthony, J. C. (2003). Possible age-associated bias in reporting of clinical features of drug dependence: Epidemiological evidence on adolescent-onset marijuana use. *Addiction*, 98(1), 71-82.

PURPOSE/OVERVIEW: As of 2003, the latest drug use reports showed a higher level of marijuana dependence for adolescents than for adults. This paper explored potential age-related differences in marijuana dependence using multivariate analysis and item-response biases.

METHODS: Marijuana dependence was measured using seven binary survey items from the 1995 to 1998 National Household Survey on Drug Abuse (NHSDA) questionnaires. Of the 86,021 respondents for these 4 years of the NHSDA, 2,628 (1,866 adolescents and 762 adults) were identified as recent-onset marijuana users. Multivariate response analysis using a generalized linear model (GLM) and generalized estimating equations (GEE) was performed to measure the age-related difference in reported marijuana use while taking into account the interdependencies of the yes-no responses and controlling for covariates. Further analyses were performed using the multiple indicators/multiple causes (MIMIC) multivariate response model to measure age-associated response bias.

RESULTS/CONCLUSIONS: The primary findings were that of the recent-onset marijuana users, younger users reported more drug dependency than older users. This association was found to be statistically significant in the multivariate analysis model, which controlled for other covariates. The MIMIC model found that there also were age-related biases in reporting of drug use. Younger users were biased toward reporting dependent behaviors. These findings support the notion that there may be differences in what constitutes "marijuana dependence" for adolescents compared with adults.

### **The utility of debriefing questions in a household survey on drug abuse**

CITATION: Fendrich, M., Wislar, J. S., & Johnson, T. P. (2003). The utility of debriefing questions in a household survey on drug abuse. *Journal of Drug Issues*, 33(2), 267-284.

PURPOSE/OVERVIEW: Respondent's discomfort in revealing answers to sensitive survey questions is often a cause of underreporting. This paper measures the effect that respondents' comfort level has on their answers to sensitive questions and whether there are any differences in mode.

METHODS: Respondents were randomly assigned to receive either the control or experimental condition. The control condition replicated the design procedures from the 1995 National Household Survey on Drug Abuse (NHSDA), which used an interviewer-administered

paper-and-pencil interviewing (PAPI) survey. The experimental condition had the choice of completing a computer-assisted personal interviewing (CAPI) survey or an audio computer-assisted self-interviewing (ACASI) survey. Following the interview, respondents were asked for a hair sample. At the end of the interview, interviewers administered two types of debriefing questions, subjective and projective, to elicit respondents' level of discomfort during the study.

**RESULTS/CONCLUSIONS:** Respondents' willingness to disclose sensitive information was correlated with their subjective and projective levels of discomfort with the survey. Contrary to the hypothesis, respondents who disclosed information revealed less comfort with the survey on the subjective debriefing items and more comfort on the projective debriefing questions. In addition, the degree of discomfort for survey respondents was greater for those in the experimental condition than the control condition. Levels of projective and subjective discomfort differed by subgroups for race, age, and education.

### **Substance use treatment need among older adults in 2020: The impact of the aging baby-boom cohort**

**CITATION:** Gfroerer, J., Penne, M., Pemberton, M., & Folsom, R. (2003). Substance use treatment need among older adults in 2020: The impact of the aging baby-boom cohort. *Drug and Alcohol Dependence*, 69, 127-135.

**PURPOSE/OVERVIEW:** The purpose of this paper is to provide estimates of the number of older adults (defined as those 50 or older) needing treatment for substance use problems in the future as the U.S. baby-boom population ages.

**METHODS:** Data from the 2000 and 2001 National Household Surveys on Drug Abuse (NHSDAs) were used to estimate logistic regression models among adults aged 50 or older, predicting treatment need in the past 12 months, which was defined as being classified with substance dependence or abuse based on DSM-IV criteria. Separate regression models were estimated for individuals depending on whether they had used alcohol by age 30. For those who had not used alcohol by age 30, age was the only predictor. For those who had used alcohol prior to age 31, predictors in the regressions included age, gender, race/ethnicity, and substance use prior to age 31. Estimated parameters from these models then were applied to a pooled sample of adults 30 or older from the 2000 NHSDA and 31 or older from the 2001 NHSDA to produce predicted probabilities of treatment need. Weights were adjusted to match census population projections on age, race, and gender. An additional adjustment was made for expected mortality rates. Standard errors for the 2020 projections were computed using jackknife replication methods.

**RESULTS/CONCLUSIONS:** The estimated number of older adults needing treatment for a substance use problem is projected to increase from 1.7 million in 2000/2001 to 4.4 million in 2020. This increase is the result of a 50 percent increase in the population aged 50 or older (from 74.8 million to 112.5 million) combined with a 70 percent increase in the rate of treatment need (from 2.3 to 3.9 percent) in the older population. Increases are projected for all gender, race, and age groups. About half of the projected 2020 population needing treatment will be aged 50 to 59, and two thirds will be male.

## **Comparison of linear and nonlinear models for generalized variance functions for the National Survey on Drug Use and Health**

CITATION: Gordek, H., & Singh, A. C. (2003). Comparison of linear and nonlinear models for generalized variance functions for the National Survey on Drug Use and Health. In *Proceedings of the 2003 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, San Francisco, CA* [CD-ROM]. Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: The generalized variance function (GVF) provides an approximation to the variance of an arbitrary domain estimate by means of modeling the relationship between the estimated variances of a selected set of domain estimates and a set of covariates, rather than through direct computation. The use of GVF is important in surveys with a very large number of characteristics, where publishing limitations and the immense amount of potential subgroupings may make several variance estimates of interest unavailable directly.

METHODS: For the National Survey on Drug Use and Health (NSDUH), GVFs were obtained using linear models with the log of the relative variance as the dependent variable, as well as a somewhat modified version that ensures the resulting design effect (deff) to be at least one. The authors consider a nonlinear generalization of these models. Numerical results on comparison of various models using the 2001 NSDUH data are presented.

RESULTS/CONCLUSIONS: The authors extended the idea of constant deff for a subset of estimates to constant deff-type parameters, such as variance-odds deff, and then proposed an alternative GVF model that overcomes the limitations of existing models. Ordinary least squares was used to fit these models to the NSDUH data, and results from different models were compared. It was found that different models performed quite similarly, and there seemed no compelling reason to change from simpler, commonly used models. However, the authors concluded that it would be useful to investigate the impact of using other models, as well as the use of weighted least squares in model fitting. For instance, the authors indicated that the effect of nonlinear modeling should be investigated. More specifically, rather than formulating a linear model for the transformed point estimate of variance, one can model linearly the mean of the variance estimate after transformation (such as log or logit).

## **Screening for serious mental illness in the general population**

CITATION: Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S. L., Manderscheid, R. W., Walters, E. E., & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*, 60(2), 184-189.

PURPOSE/OVERVIEW: The Substance Abuse and Mental Health Services Administration (SAMHSA) was charged with the task of defining "serious mental illness" (SMI) in adults and establishing a method that States can use to estimate its prevalence.

**METHODS:** This paper compared three sets of screening scales used to estimate the prevalence of SMI: the World Health Organization (WHO) Composite International Diagnostic Interview Short Form (CIDI-SF) scale, the K10 and K6 psychological distress scales, and the WHO Disability Assessment Schedule (WHODAS). A convenience sample of 155 respondents received the three screening scales using computer-assisted self-interviewing (CASI). After the self-administered questions, respondents completed the 12-month Structured Clinical Interview (SCID), including the Global Assessment of Functioning (GAF). SMI was defined in the SCID as any 12-month DSM-IV disorder (defined in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders*), besides substance use, that also had a GAF score less than 60.

**RESULTS/CONCLUSIONS:** All screening scales were found to be significantly correlated with SMI. The shortest scale, the K6, was the most statistically significant predictor of SMI. These results support the notion that short, fully developed, and carefully constructed screening scales can be strong predictors of the same results found in more lengthy and expensive clinical interviews. Another advantage of the K6 and K10 scales is that they can be administered in less time than 2 or 3 minutes, respectively.

### **The effect of interviewer experience on the interview process in the National Survey on Drug Use and Health**

**CITATION:** Odom, D. M., Eyerman, J., Chromy, J. R., McNeeley, M. E., & Hughes, A. L. (2003). The effect of interviewer experience on the interview process in the National Survey on Drug Use and Health. In *Proceedings of the 2003 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, San Francisco, CA* [CD-ROM]. Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** Analysis of survey data from the National Survey on Drug Use and Health (NSDUH) has shown a relationship between field interviewer (FI) experience, response rates, and the prevalence of self-reported substance use (Eyerman, Odom, Wu, & Butler, 2002; Hughes, Chromy, Giacoletti, & Odom, 2001, 2002; Substance Abuse and Mental Health Services Administration [SAMHSA], 2000). These analyses have shown a significant and positive relationship between the amount of prior experience an FI has with collecting NSDUH data and the response rates that a FI produces with his or her workload. These analyses also have shown a significant and negative relationship between the amount of prior experience of an FI and the prevalence of substance use reported in cases completed by that FI. In general, these analyses have been consistent with the published literature that FIs can influence both the success of the data collection process and accuracy of the population estimates (Martin & Beerteen, 1999; Singer, Frankel, & Glassman, 1983; Stevens & Bailar, 1976). Previous NSDUH analyses examined response rates and prevalence estimates independently. This made it difficult to determine whether the lower prevalence estimates for experienced FIs were a result of the change in the sample composition due to higher response rates or whether the lower prevalence estimates were a result of a direct effect of FI behavior on respondent self-reporting.

**METHODS:** This analysis combines these two explanations to produce a conceptual model that summarizes the authors' expectations for the relationship between FI experience and prevalence estimates. The combined explanation from the conceptual model is evaluated in a series of

conditional models to examine the indirect effect of response rates and the direct effect of FI experience on prevalence estimates. Earlier analyses using NSDUH data have shown a negative correlation between interviewer experience and substance use rates. However, these studies only examined the last step in the interviewing process, administering the questionnaire. This paper further explores the effect of interviewer experience by investigating a series of separate logistic models that are conditionally based on each step of the screening and interviewing (S&I) process. The S&I steps examined are contacting the household, gaining household cooperation, contacting the selected person(s), interviewing the selected person(s), and reporting of substance use. By separating the analysis into these steps, estimating the effect of interviewer experience on data collection at each stage of the survey is possible.

**RESULTS/CONCLUSIONS:** In conclusion, the analysis shows that increased FI experience simultaneously increases response rates and decreases prevalence estimates. In addition, the effect of increased FI experience on prevalence estimates cannot be fully explained by the adjustments based on earlier models (i.e., S&I level) to the final prevalence estimate model. In other words, the FI effect on prevalence cannot be fully attributed to the increase in response rates by experienced FIs. Furthermore, FI experience was significant in the final model, showing that the covariates also did not account for all the decrease in prevalence. Three hypotheses were given as possible explanations for the decrease in prevalence. As was shown in the statistical analysis, the marginal rates were too extreme to support the first hypothesis. This means that although some level of selection bias may be occurring, it is not the only cause of the decrease in prevalence estimates for experienced FIs. More likely, the relationship between FI experience and prevalence estimates is captured in hypothesis 3 (i.e., the decrease in prevalence estimates for experienced FIs is a function of lower substance use reporting by the additional respondents they obtain and also the remaining respondents that FIs with all levels of experience interview).

## **Appendix C: NSDUH changes and their impact on trend measurement**

**CITATION:** Office of Applied Studies. (2003). Appendix C: NSDUH changes and their impact on trend measurement. In *Results from the 2002 National Survey on Drug Use and Health: National findings* (HHS Publication No. SMA 03-3836, NSDUH Series H-22, pp. 107-137). Rockville, MD: Substance Abuse and Mental Health Services Administration.

**PURPOSE/OVERVIEW:** This appendix presents the results of analyses designed to determine the degree to which increases in the rates of substance use, dependence, abuse, and serious mental illness (SMI) between the 2001 National Household Survey on Drug Abuse (NHSDA) and the 2002 National Survey on Drug Use and Health (NSDUH) could be attributed to important methodological differences between the two surveys. Major changes between the 2 years included (1) a change in the survey's name, (2) providing a \$30 incentive to all interview respondents, (3) improved data collection quality control procedures, and (4) switching to the use of 2000 census data as the basis for population weighting adjustments for the 2002 survey.

**METHODS:** Six types of analyses were used to assess the degree to which methodological changes could account for the various increases in prevalence rates: (1) a retrospective cohort analysis that examined the degree to which increases in lifetime use could be attributed to new initiates; (2) a response rate pattern analysis that examined changes in response rates between the

fourth quarter of 2001 and the first quarter of 2002 for various geographic and demographic groups, reasons for refusal, and field interviewer characteristics; (3) a response rate impact analysis that attempted to determine whether the increased response rate between 2001 and 2002 resulted in "additional" respondents who in turn accounted for higher rates of substance use; (4) an analysis of the impact of new census data in which data for the 2001 survey were reweighted using 2000-based census control totals rather than 1990-based census data; (5) analyses in which measures of substance use were regressed on variables that included indicators related to the timing and occurrence of changes in data collection quality control procedures to determine the impact of these changes; and (6) a reexamination of differences in prevalence rates from the 2001 incentive experiment by applying response propensity adjusted weights to account for response rate differences and a comparison of these differences with differences in estimates between quarter 4 of the 2001 survey and quarter 1 of the 2002 survey.

**RESULTS/CONCLUSIONS:** (1) Increases in lifetime use cannot be accounted for by new initiates in 2002 or a cohort shift. (2) Response rate increases occurred across all geographic and demographic groups, with the exception of adults aged 50 or older. (3) "Additional" respondents from the response rate increase between 2001 and 2002 cannot account for the increase in prevalence rates between 2001 and 2002. (4) Reweighting the 2001 survey by using 2000 census-based control totals had minor effects on prevalence rates and a larger impact on totals. (5) Field interventions introduced during the 2001 survey appear to have had little effect on substance use estimates. (6) Differences in estimates between quarter 1 of 2002 and quarter 4 of 2001 were generally larger than differences in estimates from the incentive experiment between incentive and no-incentive groups, suggesting that incentive effects alone do not account for overall differences.

### **Effect of incentives on data collection: A record of calls analysis of the National Survey on Drug Use and Health**

**CITATION:** Painter, D., Chromy, J., Meyer, M., Granger, R. A., & Clarke, A. (2003). Effect of incentives on data collection: A record of calls analysis of the National Survey on Drug Use and Health. In *Proceedings of the 2003 Joint Statistical Meetings, American Statistical Association, AAPOR - Section on Survey Research Methods, San Francisco, CA* (pp. 170-176). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** Given the decline in response rates during the late 1990s, a randomized, split-sample experiment was conducted during the first 6 months of data collection for the 2001 National Survey on Drug Use and Health (NSDUH) to evaluate the effectiveness of monetary incentives in improving respondent cooperation. Based on the outcome of the 2001 experiment, NSDUH interviewers began offering a \$30 incentive to all survey respondents in 2002. This paper analyzes the effect of the new \$30 incentive on the data collection process as measured by record of calls (ROC) information.

**METHODS:** An ROC was generated each time an interviewer visited a household. It consists of a code that describes the outcome of the visit along with any notes an interviewer deems pertinent to a future visit. The data contain extensive information on the amount of effort taken to contact and obtain cooperation from respondents. Using these data, the authors analyzed the

impact of the incentives on contact and cooperation patterns. The authors developed measures of the intensity of the follow-up, which include the number of calls, the timing of calls, special follow-up letters, the use of alternative interviewers, and similar special procedures used in trying to achieve contact and cooperation. These measures of the intensity of the follow-up were then related to household contact rates, household screening cooperation rates, respondent contact rates, and respondent cooperation rates. The authors first provided a brief description of the change in response rates between 2001 and 2002. Then they presented a model-based analysis assessing the changes in the patterns after the introduction of the incentives, while controlling for interviewer and respondent characteristics.

**RESULTS/CONCLUSIONS:** NSDUH data showed that the increases in response rates that accompany the use of incentives were also accompanied by the need for fewer call days and fewer calls to finalize sample dwelling units and sample persons. A review of the demographic data by call order showed that the sample distribution on demographic measures could be changed by prematurely cutting off the follow-up of pending cases. More study is needed to determine what effect such policies might have on the principal study measures on substance use addressed in NSDUH. The monitoring of ROC data provides a useful tool for ensuring that adequate follow-up procedures are being used within the limits of reasonable cost management.

*This page intentionally left blank*

## 2004

### **Triad sampling in household surveys**

CITATION: Aldworth, J., & Chromy, J. (2004). Triad sampling in household surveys. In *Proceedings of the 2004 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Toronto, Canada* [CD-ROM]. Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: The motivation for selecting three individuals from a dwelling unit (DU) derives from an interest in studying the behavioral relationships among certain triads (e.g., two parents and a child, a parent and two children). Computer-assisted DU screening provides the mechanism for targeted sampling of individuals, pairs, and even triads within a DU. Chromy and Penne (2002) showed how a modification of Brewer's (1963, 1974) method for samples of size two was used to select samples of 0, 1, or 2 individuals from eligible DUs. They also developed a second adaptation to control the number of pairs selected.

METHODS: Chromy and Penne's modification of Brewer's method and their adaptation to control the number of individuals selected from a DU is extended to the case of sampling triads within DUs. Some empirical data on household roster composition, and response rates based on the number of individuals selected, are presented from the National Survey on Drug Use and Health (NSDUH).

RESULTS/CONCLUSIONS: The results of simulations of the sample selection and response process are presented as a means of evaluating alternatives. Possible negative impacts of triad selections on design effect, response rates, and response bias were assessed. The results indicate that although pair selections in the 2002 NSDUH show little negative impact, a field test assessment is needed for triad assessment.

### **Estimating substance abuse treatment need by state [editorial]**

CITATION: Gfroerer, J., Epstein, J., & Wright, D. (2004). Estimating substance abuse treatment need by state [editorial]. *Addiction*, 99(8), 938-939.

PURPOSE/OVERVIEW: This editorial comments on the methods used by the Substance Abuse and Mental Health Services Administration (SAMHSA) to measure the need for the treatment of substance use disorders at the State level.

METHODS: The two methods compared are the one used in the National Survey on Drug Use and Health (NSDUH) and the index method proposed by W. E. McAuliffe and R. Dunn in this issue of *Addiction*. The authors describe the strengths and weaknesses of each of these methods.

RESULTS/CONCLUSIONS: The strengths of NSDUH are that it uses an independent probability sample for each State, data collection procedures and methods are defined and carried out consistently across States, and the data are current, reflecting well-defined time periods.

A validation study also has demonstrated that NSDUH has small biases and allows for estimates of different subgroups. The main limitations are the small sample sizes in each State and possible underreporting and undercoverage. The primary weakness of the index method is that their measures are not well defined and therefore not consistently interpreted. For these reasons, the NSDUH is recognized as more valid than the index method, but the NSDUH results are meant to be interpreted along with the results from other data sources.

### **Estimating trends in substance use based on reports of prior use in a cross-sectional survey**

CITATION: Gfroerer, J., Hughes, A., Chromy, J., Heller, D., & Packer, L. (2004). Estimating trends in substance use based on reports of prior use in a cross-sectional survey. In S. B. Cohen & J. M. Lepkowski (Eds.), *Eighth Conference on Health Survey Research Methods* (HHS Publication No. PHS 04-1013, pp. 29-34). Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

PURPOSE/OVERVIEW: Substance use trends in the United States have shown dramatic shifts since the 1960s. Among youths aged 12 to 17, the rate of past month marijuana use was less than 2 percent in the early 1960s, increased to 14 percent by 1979, then decreased to 3.4 percent in 1992 before rising to 8.2 percent in 1995. Major shifts in prevalence at different points in time and for different age groups have been observed for other substances, including cocaine, LSD, Ecstasy, opiates, cigars, and cigarettes (Substance Abuse and Mental Health Services Administration [SAMHSA], 2003). Accurate measurement of these trends is critical for policymakers targeting limited resources efficiently toward emerging problems. Trend data also are used for assessing the impact of prevention and treatment programs. The typical method used for measuring substance use trends is comparing prevalence estimates across repeated cross-sectional surveys. An alternative approach is to collect data about prior substance use within a cross-sectional survey and to construct prevalence estimates for prior years based on these data. Besides the cost advantages, these retrospective estimates have some analytic advantages. When data are obtained for different periods from the same respondents, trend analyses are more powerful, due to the positive correlation between estimates, as is the case in a longitudinal study. Retrospective estimates also may be the only alternative if estimates are needed for periods for which direct estimates are not available. However, retrospective estimates do have important limitations. Bias due to recall decay, telescoping, and reluctance to admit socially undesirable behaviors could cause underestimation or distort trends (Johnson, Gerstein, & Rasinski, 1998; Kenkel, Lillard, & Mathios, 2003). Bias also could result from coverage errors affecting the capability of the sample to represent the population of interest for prior time periods, due to mortality, immigration, or other changes in the population. This paper discusses several types of retrospective estimates and presents analyses of data from the National Survey on Drug Use and Health (NSDUH) to assess biases in these estimates.

METHODS: Several analyses were undertaken to assess bias in retrospective estimates. One known source of bias in retrospective estimates is the inclusion of data from immigrants who were not living in the United States in some prior years. The authors compared estimates of incidence and lifetime use (for those aged 12 to 17 or 18 to 25) for the full 2002 NSDUH sample

with estimates based on the sample excluding these immigrants, according to questions on country of birth and years in the United States.

Trends in incidence estimates for 1965 to 1990 based on 1991 to 1993 data (shortest recall), 1994 to 1998 data, 1999 to 2001 data, and 2002 data (longest recall) were compared. For the 1991 to 1997 period, trends based on the 1994 to 1998 data, 1999 to 2001 data, and 2002 data were compared. Consistency was assessed through visual inspection of curves and with correlations. Because of methodology changes, comparisons of levels from different sets of surveys were not made.

The authors compared 2002-based retrospective lifetime use estimates (excluding immigrants) with direct lifetime use estimates from earlier NSDUHs (for those aged 12 to 17 or 18 to 25) and from the Monitoring the Future (MTF) study, a survey of high school seniors (Johnston, O'Malley, & Bachman, 2003). To reduce the effect of sampling error, the authors combined several years of data, depending on availability, and generated average annual lifetime prevalence for specific time periods. Because 1999 and 2002 survey changes resulted in increased reporting of lifetime use, the authors expected retrospective estimates to be greater than the direct estimates for years before 1999.

The authors compared retrospective lifetime use estimates for 2002, based on 2003 NSDUH data (first 6 months of data currently available), to direct 2002 lifetime use estimates, from the 2002 NSDUH (first 6 months of data, for consistency). Comparisons were made for 19 substances for those aged 12 to 17 or 18 to 25.

To assess the accuracy of these estimates, the authors compared January to June 2003-based retrospective estimates of past year use in January to June 2002 to direct past year estimates from the January to June 2002 data, by age group.

**RESULTS/CONCLUSIONS:** Marijuana incidence estimates for 1965 to 2001 were 2.5 percent higher when immigrants were included. For most other illicit drugs, the bias was smaller, indicating that very little initiation for these drugs occurs among immigrants prior to their entry to the United States. However, biases for alcohol and cigarette incidence estimates were larger (8 percent for alcohol, 7 percent for cigarettes). In general, they were largest for the years 1979 to 1994 (3.5 percent for marijuana, 11 percent for alcohol, 10 percent for cigarettes) and smallest for years after 1997 (1 percent for marijuana, 3 percent for alcohol, and 3 percent for cigarettes). For lifetime prevalence rates, bias due to including immigrants was negative for nearly every substance because of the low rates of substance use among immigrants. For youth estimates during the period from 1979 to 1990, the inclusion of immigrants resulted in biases of about -14 percent for marijuana, -15 percent for cocaine, -9 percent for cigarettes, and -9 percent for alcohol. Bias was generally worse for estimates for those aged 12 to 17 than for those aged 18 to 25, and there was very little bias in any estimates for years after 1997. Estimates of alcohol use for those aged 18 to 25 including immigrants showed very small but positive bias (1.5 percent) for the period from 1982 to 1993 and a larger positive bias (7 percent) for 1965 to 1981.

## **A simple evaluation of the imputation procedures used in NSDUH**

**CITATION:** Grau, E., Frechtel, P., Odom, D., & Painter, D. (2004). A simple evaluation of the imputation procedures used in NSDUH. In *Proceedings of the 2004 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Toronto, Ontario, Canada* [CD-ROM]. Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** The National Survey on Drug Use and Health (NSDUH) is the primary source of information on drug use in the United States. Since 1999, the predictive mean neighborhood (PMN) procedure has been used to impute missing values for many of the analytical variables. This method is a combination of two commonly used imputation methods: a nearest neighbor hot-deck and a modification of Rubin's predictive mean matching method. Although PMN has many practical advantages, it has not been evaluated formally.

**METHODS:** The authors proposed a simple simulation to evaluate PMN. Using only complete data cases, they induced random patterns of missingness in the data for selected outcome variables. Imputations then were conducted using PMN and a weighted nearest neighbor hot deck. This process of inducing missingness and imputing missing values was repeated multiple times. The imputed values using PMN and the weighted hot deck then were compared with the true values found in the complete data across the repeated iterations. In particular, the authors compared the number of matches between the two methods, as well as compared statistics derived from the data, such as drug prevalence estimates.

**RESULTS/CONCLUSIONS:** This evaluation showed that using PMN provided a modest advantage in the match test, and, in the missing completely at random (MCAR) case, using weighted sequential hot deck (WSHD) provided a modest advantage in the mean test. In each of these cases, however, the degree of difference between the methods, though significant, was not substantial. Not surprisingly, all methods did badly in the mean test when the data were not missing at random (NMAR).

## **Measuring treatment needs: A reply to Gfroerer, Epstein, and Wright**

**CITATION:** McAuliffe, W. E. (2004). Measuring treatment needs: A reply to Gfroerer, Epstein, and Wright. *Addiction*, 99(9), 1219-1220.

**PURPOSE/OVERVIEW:** This commentary is a reply to the editorial titled "Estimating Substance Abuse Treatment Need by State" by Joe Gfroerer, Joan Epstein, and Doug Wright. The author commented on the strengths and weaknesses of studies used to measure the treatment need for substance use disorders.

**METHODS:** The two studies compared were the National Survey on Drug Use and Health (NSDUH) and the Substance Abuse Need Index (SNI) produced by W. E. McAuliffe and R. Dunn (2004). The author discussed the strengths and weaknesses of each approach.

**RESULTS/CONCLUSIONS:** NSDUH suffers from underreporting, undercoverage, and nonresponse, which likely leads to undercounting users of hard-core drugs, such as cocaine.

The SNI more accurately captures the incarcerated and homeless populations, improving their estimates for cocaine use and drug abuse. However, the NSDUH estimates probably are more accurate for marijuana use. The author disagreed with Gfroerer et al. that the SNI was less reliable or valid than NSDUH and expressed the opinion that multiple data sources need to be considered together to get the full picture on drug abuse and need for treatment.

### **Substance abuse treatment needs and access in the USA: Interstate variations**

CITATION: McAuliffe, W. E., & Dunn, R. (2004). Substance abuse treatment needs and access in the USA: Interstate variations. *Addiction*, 99(8), 999-1014.

PURPOSE/OVERVIEW: This paper analyzed two measures for substance abuse treatment needs across States. They were used to explore geographic variations in substance abuse, the causes for substance abuse, the stability of these estimates over time, and whether the severity of substance abuse was correlated to need.

METHODS: This study used alcohol dependency estimates from the National Survey on Drug Use and Health (NSDUH), and drug and alcohol use indexes based on alcohol-related mortality and arrests data, to measure interstate differences between substance abuse treatment needs and treatment services. This study tested the reliability and the validity of the survey measures used. The index for substance abuse treatment then was regressed on the measures for substance abuse need to identify differences in availability of treatment across States.

RESULTS/CONCLUSIONS: The individual indicators of treatment needs and availability of treatment for substance abuse across the United States had reliability and construct validity. Substance use problems in the United States are clustered geographically, with the most severe problems appearing in the western States. The biggest gaps between treatment need and treatment access appeared in the South where there was moderate need for treatment, but very low access. The interstate discrepancies in treatment need versus treatment services indicated that substance use problems in rural areas are often overlooked by treatment services.

### **Prevalence of adult binge drinking: A comparison of two national surveys**

CITATION: Miller, J. W., Gfroerer, J. C., Brewer, R. D., Naimi, T. S., Mokdad, A., & Giles, W. H. (2004). Prevalence of adult binge drinking: A comparison of two national surveys. *American Journal of Preventive Medicine*, 27(3), 197-204.

PURPOSE/OVERVIEW: This paper compares two national surveys—the Behavioral Risk Factor Surveillance System (BRFSS) and the National Survey on Drug Use and Health (NSDUH)—that measure the prevalence of binge drinking across States and overall for the United States. The authors examine methodological differences between the two studies to assess their impact on the survey estimates.

METHODS: The main methodological difference between the two studies is that BRFSS is conducted over the telephone and NSDUH is conducted in-person using audio computer-assisted self-interviewing (ACASI). BRFSS assesses binge drinking using three alcohol questions, and

NSDUH uses nine; however, the questions are very similar. Because BRFSS uses a telephone sample and NSDUH is an in-person survey, data from NSDUH were limited to only telephone-households to make comparisons consistent. In addition, NSDUH was restricted to respondents aged 18 years or older to match BRFSS. Response rates, survey size, and other characteristics of the two studies were compared.

**RESULTS/CONCLUSIONS:** BRFSS binge drinking estimates for the United States and most States were considerably lower than the NSDUH estimates, even after stratifying for sociodemographic variables. The demographic characteristics of the sample were very similar; the majority were male and white, non-Hispanic. However, there were no significant differences for binge drinking in the past 30 days between the two studies. The large differences in estimates was likely due to the use of ACASI in NSDUH, which is perceived as more anonymous and yields higher reporting of sensitive behaviors. Other possible explanations for the differences are the sample size, the number of questions asked about alcohol, and the overall topic of the survey.

### **A system for detecting interviewer falsification**

**CITATION:** Murphy, J., Baxter, R. K., Eyerman, J., Cunningham, D., & Barker, P. (2004). A system for detecting interviewer falsification. In *Proceedings of the 2004 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Toronto, Ontario, Canada* [CD-ROM]. Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** The National Survey on Drug Use and Health (NSDUH) has developed a detailed system for the detection of interviewer falsification. This process includes phone, mail, and in-person field verification procedures, as well as the review of interview and interviewer-level process data to identify cases and interviewers requiring extra verification efforts. Although these components of the system successfully identify a majority of potential falsifiers, more savvy falsifiers may be able to remain undetected if they are aware that their process data are being scrutinized. To address this gap, NSDUH added a new component to the falsification detection system: the regular review of interview response and question-level timing data. This paper details the structure and operationalization of this system and presents examples of its effectiveness on NSDUH.

**METHODS:** Based on what is known about the area in which an interviewer is working and the types of cases he or she is assigned, a likely range of interview responses is calculated. Response distributions are compared with this likely range at the interviewer level to identify interviewers whose responses appear to be highly unlikely, given their caseloads. These additional measures make it even more difficult for falsifiers to remain undetected because they would need to have a specific understanding of the prevalence and correlates of substance use in order to enter likely responses. Similarly, question-level timings for particular items that require certain interviewer-responder interactions are compared with a "gold standard" to detect outliers. Once potential falsifiers are identified, the work of the suspected interviewers is subject to 100 percent and/or in-person verification.

**RESULTS/CONCLUSIONS:** The analyses in this paper showed that falsification detection can be improved through the systematic review of response data and metadata, such as module and

item timings. Through early detection and remediation, the threat of falsification to survey bias and increased costs can be reduced. Although the NSDUH system has improved based on these recent enhancements, there are still many more enhancements that could be incorporated. In particular, the authors planned to assess the following techniques for possible adoption: (1) analysis of screening data, (2) analysis of record of calls data and other metadata, (3) a data mining approach, and (4) statistical process control.

## **Nonresponse among persons age 50 and older in the National Survey on Drug Use and Health**

**CITATION:** Murphy, J., Eyerman, J., & Kennet, J. (2004). Nonresponse among persons age 50 and older in the National Survey on Drug Use and Health. In S. B. Cohen & J. M. Lepkowski (Eds.), *Eighth Conference on Health Survey Research Methods* (HHS Publication No. PHS 04-1013, pp. 73-78). Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

**PURPOSE/OVERVIEW:** Response rates traditionally have been highest among the youngest respondents and lowest among the oldest, with the lowest rates found in the 50 or older (50+) age group. The introduction in 2002 of a series of methodological enhancements to the National Survey on Drug Use and Health (NSDUH) appeared to improve the response rates for most age groups but had only a small impact on the 50+ age group (Kennet, Gfroerer, Bowman, Martin, & Cunningham, 2003). Because lower response rates make nonresponse bias more likely, and because there is a disturbingly low response rate among the 50+ group, this paper aims to understand why this may be in order to understand how the problem can be ameliorated. This topic is of increasing importance as the proportion of Americans in this age group increases (U.S. Bureau of the Census, 1999). Obtaining unbiased survey estimates will be vital to assess accurately the substance abuse treatment need for older Americans in the coming years. This need is expected to nearly triple by 2020 as the baby boom generation carries its alcohol and drug use into older ages (Gfroerer, Penne, Pemberton, & Folsom, 2002). The purpose of this paper is to provide a better understanding of nonresponse among older sample members in NSDUH in order to tailor methods to improve response rates and reduce the threat of nonresponse error.

**METHODS:** This paper examines the components of nonresponse (refusals, noncontacts, and/or other incompletes) among the 50+ age group in NSDUH. It also examines respondent, environmental, and interviewer characteristics in order to identify the correlates of nonresponse among the 50+ group, including relationships that are unique to the 50+ group. Finally, this paper considers the root causes for differential nonresponse by age, drawing from focus group sessions with NSDUH field interviewers on the topic of nonresponse among the 50+ group.

**RESULTS/CONCLUSIONS:** This paper showed that nonresponse in NSDUH was higher among the 50+ group than among any other age group and was primarily due to a high rate of refusals, especially among sample members aged 50 to 69, and a high rate of physical and mental incapability among those 70 or older. Taken together with evidence from interviewer focus groups, it appeared that the higher rate of refusal among the 50+ age group may, in part, have

been due to fears and misperceptions about the survey and interviewers' intentions. Increased public awareness about the study may allay these fears. Although an increase in the incentive amount may not automatically increase response rates among this group, the authors concluded that other protocol changes and methodological enhancements may be effective.

### **Imputation and unbiased estimation: Use of the centered predictive mean neighborhoods method**

CITATION: Singh, A., Grau, E., & Folsom, R. (2004). Imputation and unbiased estimation: Use of the centered predictive mean neighborhoods method. In *Proceedings of the 2004 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Toronto, Ontario, Canada* [CD-ROM]. Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: Methods for determining the predictive distribution for multivariate imputation range between two extremes, both of which are commonly employed in practice: a completely parametric model-based approach, and a completely nonparametric approach, such as the nearest neighbor hot deck (NNHD). A semiparametric middle ground between these two extremes is to fit a series of univariate models and construct a neighborhood based on the vector of predictive means. This is what is done under the predictive mean neighborhood (PMN) method, a generalization of Rubin's predictive mean matching method. Because the distribution of donors in the PMN neighborhood may not be centered at the recipient's predictive mean, estimators of population means and totals could be biased. To overcome this problem, the authors propose a modification to PMN that uses sampling weight calibration techniques, such as the generalized exponential model (GEM) method of Folsom and Singh to center the empirical distribution from the neighborhood.

METHODS: Empirical results on bias and mean squared error (MSE), based on a simulation study using data from the 2002 National Survey on Drug Use and Health, are presented to compare the centered PMN with other methods.

RESULTS/CONCLUSIONS: Although it had been theorized that bias could be a problem for existing methods, this simulation study was unable to detect any meaningful bias with any of the methods. Furthermore, none of the methods showed a consistent pattern of higher or lower variance beyond what was expected.

### **Combined-year state-level public use files and single-year nation-level PUFs from the National Survey of [sic] Drug Use and Health (NSDUH) data**

CITATION: Wright, D., & Singh, A. (2004). Combined-year state-level public use files and single-year nation-level PUFs from the National Survey of [sic] Drug Use and Health (NSDUH) data. In *Proceedings of the 2004 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Toronto, Ontario, Canada* [CD-ROM]. Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** Since 1999, the Substance Abuse and Mental Health Services Administration (SAMHSA) has provided yearly national public use files (PUFs) for the National Survey on Drug Use and Health (NSDUH) data using a procedure based on the micro agglomeration, substitution, subsampling, and calibration (MASSC) system for statistical disclosure limitation. There is a growing demand for State-level data, and SAMHSA is considering providing State-level PUFs based on combining several years of NSDUH data. The authors explore various concerns and approaches to State-level PUFs and indicated how MASSC could address some of them.

**METHODS:** Releasing combined-year State-level PUFs alongside single-year national PUFs poses several challenges. The most important one is that confidentiality of an individual could be compromised if an intruder were able to match the State-level PUFs with the national PUFs on the basis of various sensitive variables that are typically not perturbed, and thus may succeed in attaching State identifiers to the national PUFs. This problem can be reduced by taking advantage of the randomness in perturbation and suppression used in MASSC.

**RESULTS/CONCLUSIONS:** In this paper, the authors suggest a way in which State-level PUFs could be created if one is willing to combine data over several years. The State PUFs would provide a wealth of information for each State for the calculation of point estimates and for analyzing relationships within each State.

*This page intentionally left blank*

## 2005

### **A test of the item count methodology for estimating cocaine use prevalence**

CITATION: Biemer, P. P., Jordan, B. K., Hubbard, M., & Wright, D. (2005). A test of the item count methodology for estimating cocaine use prevalence. In J. Kennet & J. Gfroerer (Eds.), *Evaluating and improving methods used in the National Survey on Drug Use and Health* (HHS Publication No. SMA 05-4044, Methodology Series M-5, pp. 149-174). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

PURPOSE/OVERVIEW: The Substance Abuse and Mental Health Services Administration (SAMHSA) has long sought ways to improve the accuracy of the prevalence estimates provided by the National Survey on Drug Use and Health (NSDUH). One method of data collection that shows some promise for improving reporting accuracy is the "item count method." This technique provides respondents with an enhanced perception of anonymity when reporting a sensitive behavior, such as drug use. Past experience with the item count (IC) methodology (e.g., Droitcour et al., 1991) has identified two major problems with this method: task difficulty and the selection of the innocuous items for the list.

METHODS: To test the efficacy of the IC methodology for estimating drug use prevalence, the method was implemented in the 2001 survey. This chapter describes the research conducted in 2000 and 2001 to (1) develop an IC module for past year cocaine use, (2) evaluate and refine the module using cognitive laboratory methods, (3) develop the prevalence estimators of past year cocaine use for the survey design, and (4) make final recommendations on the viability of using the IC method to estimate drug use prevalence. As part of item (4), IC estimates of past year cocaine use based on this implementation are presented, and the validity of the estimates is discussed.

RESULTS/CONCLUSIONS: Considerable effort was directed toward the development, implementation, and analysis of an IC methodology for the estimation of cocaine use prevalence in NSDUH. Several adaptations of existing methods were implemented, offering hope that the refined method would succeed in improving the accuracy of prevalence estimation. Despite these efforts, the IC methodology failed to produce estimates of cocaine use that were even at the level of those obtained by simply asking respondents directly about their cocaine use. Because the direct questioning method is believed to produce underestimates of cocaine, these findings suggest that the IC methodology is even more biased than self-reports.

### **Model-based estimation of drug use prevalence using item count data**

CITATION: Biemer, P., & Brown, G. (2005). Model-based estimation of drug use prevalence using item count data. *Journal of Official Statistics*, 21(2), 287-308.

PURPOSE/OVERVIEW: The item count (IC) method for estimating the prevalence of sensitive behaviors was applied to the National Survey on Drug Use and Health (NSDUH) to estimate the

prevalence of past year cocaine use. In spite of the considerable effort and research to refine and adapt the IC method to this survey, the method failed to produce estimates that were larger than the estimates based on self-reports. There is evidence to indicate that this problem could be a measurement error in the IC responses.

**METHODS:** A new model-based estimator was proposed to correct the IC estimates for measurement error, and it was found that the new estimator produced less biased prevalence estimates. The model combined the IC data, replicated measurements of the IC items, and used responses to the cocaine use question in order to obtain estimates of the classification error in the observed data. The data were treated as fallible indicators of latent true values. To obtain an identifiable model, traditional latent class analysis assumptions were made.

**RESULTS/CONCLUSIONS:** The resulting estimates of cocaine use prevalence were approximately 43 percent larger compared with estimates based only on self-report. The estimated underreporting rates were also consistent with those estimated from other studies of drug use underreporting.

## **Evaluation of follow-up probes to reduce item nonresponse in NSDUH**

**CITATION:** Caspar, R. A., Penne, M. A., & Dean, E. (2005). Evaluation of follow-up probes to reduce item nonresponse in NSDUH. In J. Kennet & J. Gfroerer (Eds.), *Evaluating and improving methods used in the National Survey on Drug Use and Health* (HHS Publication No. SMA 05-4044, Methodology Series M-5, pp. 121-148). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** Technological advances in survey research since 1990 offer the ability to improve the quality of face-to-face survey data in part by reducing item nonresponse. Assuming the instrument has been programmed correctly, computer-assisted interviewing (CAI) can eliminate missing data caused by confusing skip instructions, hard-to-locate answer spaces, and simple inattention to the task at hand. Yet in and of itself, CAI cannot reduce the item nonresponse created when a respondent chooses to give a "Don't know" or "Refused" response to a survey question. Previous research (see, e.g., Turner, Lessler, & Gfroerer, 1992) has shown that these types of responses are more common in self-administered questionnaires than in those administered by an interviewer. Most likely, this occurs because in a self-administered interview the interviewer does not see the respondent's answer and thus cannot probe or follow up on these types of incomplete responses. This chapter introduces a methodology designed to reduce item nonresponse to critical items in the audio computer-assisted self-interviewing (ACASI) portion of the questionnaire used in the National Survey on Drug Use and Health (NSDUH). Respondents providing "Don't know" or "Refused" responses to items designated as essential to the study's objectives received tailored follow-up questions designed to simulate interviewer probes.

**METHOD:** The results are based on an analysis of unweighted data collected for the 2000 survey ( $n = 71,764$ ). In total, 2,122 respondents (3.0 percent) triggered at least one of the 38-item nonresponse follow-up questions in 2000. Demographic characteristics are provided of those respondents who triggered at least one follow-up item. To determine what other respondent

characteristics tend to be associated with triggering follow-up questions, several multivariate models were developed. Logistic regression was used to determine the likelihood of triggering a follow-up (e.g., answering "Don't know" or "Refused" to a critical lifetime, recency, or frequency drug question). The lifetime follow-up model was run on all cases, while the recency and frequency follow-up models were run only on those subsets of respondents who reported lifetime use (in the case of the recency follow-up item) or who reported use in the past 30 days (in the case of the frequency follow-up). The predictor variables included in these models are described.

**RESULTS/CONCLUSIONS:** Perhaps the most significant finding from these analyses is that item nonresponse to these critical items was quite low. For the most part, respondents were willing to answer these questions and did not require additional prompting to do so. As a result of the low item nonresponse rates, the data presented must be interpreted with care. The results of these analyses suggest that younger respondents were more likely to trigger the follow-ups and to provide substantive responses to the follow-ups. In addition, the follow-up methodology was more successful in converting respondents who triggered the follow-up through a "Don't know" response than through a "Refused" response. The methodology also was more successful when combined with a revised question that reduced respondent recall burden, as was done with the 30-day frequency follow-up for "Don't know." The largest percentage of follow-up responders provided a substantive response to the 30-day frequency follow-up when the question was simplified by providing response categories in place of the open-ended response field. There also was some evidence to suggest that drug use may be more prevalent among the follow-up responders although small sample sizes precluded a thorough examination of this result.

Taken together, these results suggest the follow-up methodology is a useful strategy for reducing item nonresponse, particularly when the nonresponse is due to "Don't know" responses. Additional thought should be given to whether improvements can be made to the refusal follow-ups to increase the number of respondents who convert to a substantive response. Focus groups could be useful in identifying other reasons (beyond the fear of disclosure and questions about the importance of the data) that could cause respondents to refuse these critical items. The results of such focus groups could be used to develop more appropriately worded follow-ups that might be more persuasive in convincing respondents that they should provide substantive responses.

### **Association between interviewer experience and substance use prevalence rates in NSDUH**

**CITATION:** Chromy, J. R., Eyerman, J., Odom, D., McNeeley, M. E., & Hughes, A. (2005). Association between interviewer experience and substance use prevalence rates in NSDUH. In J. Kennet & J. Gfroerer (Eds.), *Evaluating and improving methods used in the National Survey on Drug Use and Health* (HHS Publication No. SMA 05-4044, Methodology Series M-5, pp. 59-88). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** Analysis of survey data from the National Survey on Drug Use and Health (NSDUH) has shown a relationship between interviewer experience, response rates, and the prevalence of self-reported substance use (Eyerman, Odom, Wu, & Butler, 2002; Hughes,

Chromy, Giacoletti, & Odom, 2001, 2002). These analyses have shown a significant and positive relationship between the amount of prior experience an interviewer has with collecting NSDUH data and the response rates that the interviewer produces with his or her workload. The analyses also have shown a significant and negative relationship between the amount of prior experience of an interviewer and the prevalence of substance use reported in cases completed by that interviewer. This chapter describes the methodology employed to explain these effects within a unified theoretical framework.

**METHODS:** The prior analyses mentioned above examined interviewer response rates and prevalence independently. This has made it difficult to determine whether the lower prevalence estimates for experienced interviewers are a result of the change in the sample composition due to higher response rates or whether the lower prevalence estimates are a result of a direct effect of interviewer behavior on respondent self-reporting. This study combines these two explanations to produce a conceptual model that summarizes the expectations for the relationship between interviewer experience and prevalence estimates. The combined explanation from the conceptual model is evaluated in a series of conditional models to examine the indirect effect of response rates and the direct effect of interviewer experience on prevalence estimates.

**RESULTS/CONCLUSIONS:** The analysis shows that increased interviewer experience simultaneously increases response rates and decreases prevalence estimates. In addition, the effect of increased interviewer experience on prevalence cannot be fully explained by weight adjustments based on earlier models (i.e., screening and interview level). In other words, the interviewer effect on prevalence cannot be fully attributed to the increase in response rates by experienced interviewers. Furthermore, interviewer experience was significant in the final model, showing that the covariates also did not account for all the decrease in prevalence. A statistical analysis of marginal and incremental prevalence estimates based on three levels of interviewer experience showed that plausible explanations for the decrease in prevalence for experienced interviewers include (1) lower substance use reporting by the additional respondents and (2) lower reporting of substance use by respondents that interviewers with all levels of experience interview.

## **Comparing NSDUH income data with income data in other datasets**

**CITATION:** Cowell, A. J., & Mamo, D. (2005). Comparing NSDUH income data with income data in other datasets. In J. Kennet & J. Gfroerer (Eds.), *Evaluating and improving methods used in the National Survey on Drug Use and Health* (HHS Publication No. SMA 05-4044, Methodology Series M-5, pp. 175-188). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** Personal income and family or household income are among the many demographic measures obtained in the National Survey on Drug Use and Health (NSDUH). Because income is a correlate of substance use and other behaviors, it is important to evaluate the accuracy of the income measure in NSDUH. One metric of accuracy is to compare the estimated distribution of income based on NSDUH with the distributions from other data sources that are used frequently. This chapter compares the distribution of 1999 personal income data

from the 2000 NSDUH with the distributions in the same year from the Current Population Survey (CPS) and the statistics of income (SOI) data.

**METHODS:** The income measure used from the two surveys (NSDUH and the CPS) is personal income, rather than family or household income. Although the CPS explicitly gathers information about more sources of personal income than NSDUH, the income sources in the CPS were combined to map exactly to NSDUH. The income measure from the SOI is adjusted gross income (AGI), reported by various combinations of filing status. Two sets of comparisons were made that differed by marital status in the surveys and filing status in the SOI. First, the income distribution reported in the surveys, regardless of marital status, was compared with that in the SOI, regardless of filing status. For those tax-filing units whose status was "married filing jointly," the reported AGI was for two people, whereas the survey data used were for individual income only. Consequently, these comparisons were unlikely to provide a close fit of the income distributions, particularly in higher income intervals. Second, the income distribution in the surveys for those who were unmarried (excluding widow[er]s) was compared with the income distribution in the SOI for those whose filing status was single. Because only unmarried people can file as "single," the restrictions in this second set of comparisons should have ensured a relatively close fit between the survey data and the SOI. The data did not allow a reasonable comparison to be made between the income distribution in NSDUH and the distribution for other filing statuses in the SOI, such as "married, filing jointly." Because pair-level weights were not available for NSDUH at the time of this analysis, a reporting unit in NSDUH could not be created so that it could compare reasonably with "married filing jointly" in the SOI. NSDUH weights are calibrated to represent individuals. For NSDUH data to represent a pair of married people, rather than individuals, a different set of weights—pair-level weights—are required.

**RESULTS/CONCLUSIONS:** Despite some fundamental differences between the SOI and either of the survey datasets (CPS and NSDUH), there were strong similarities between the three income distributions. In both sets of comparisons, the frequencies reported in NSDUH and the CPS were typically within 2 percentage points of each other across all income intervals. With the exception of the lowest interval, in the second set of comparisons (single people and single filers), the frequencies of the three datasets were within 2.5 percentage points of one another across all income intervals.

## **Incidence and impact of controlled access situations on nonresponse**

**CITATION:** Cunningham, D., Flicker, L., Murphy, J., Aldworth, J., Myers, S., & Kennet, J. (2005). Incidence and impact of controlled access situations on nonresponse. In *JSM Proceedings, 63rd Annual Conference of the American Association for Public Opinion Research, American Statistical Association* (pp. 3841-3843). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** Failure to collect data from dwelling units with controlled access (i.e., situations where an obstacle keeps an interviewer from reaching respondents) may introduce bias through systematic underrepresentation of certain subgroups. For example, high-income or urban households are frequently found in controlled access situations compared with other subgroups in the United States. This paper summarizes the incidence of controlled access by dwelling unit

type and State for all 169,535 of the 2004 National Survey on Drug Use and Health (NSDUH) sample dwelling units and introduces a model that predicts the effects of controlled access barriers on unit and item nonresponse.

**METHODS:** The authors cross-tabulated controlled access and housing characteristics data to describe the 2004 NSDUH sample. Then they developed regression models to predict unit and item nonresponse, with the expectation that controlled access and housing units other than single family homes contribute to nonresponse.

**RESULTS/CONCLUSIONS:** The authors found that the rate of controlled access was comparable with that reported in previous studies. As predicted, housing units with some form of controlled access were less likely to be successfully screened or interviewed. In addition to discussing these findings, the authors presented ideas for further investigation of the role that controlled access barriers may have on nonresponse error and data quality.

### **The differential impact of incentives on cooperation and data collection costs: Results from the 2001 National Household Survey on Drug Abuse incentive experiment**

**CITATION:** Eyerman, J., Bowman, K., Butler, D., & Wright, D. (2005). The differential impact of incentives on cooperation and data collection costs: Results from the 2001 National Household Survey on Drug Abuse incentive experiment. *Journal for Social and Economic Measurement*, 30(2-3), 157-169.

**PURPOSE/OVERVIEW:** Research has shown that cash incentives can increase cooperation rates and response rates in surveys. The purpose of this paper is to determine whether the impact of incentives in cooperation is consistent across subgroups.

**METHODS:** An experiment on differing levels of incentives was conducted during the 2001 National Household Survey on Drug Abuse (NHSDA). Respondents were randomly assigned to receive either a \$40 incentive, a \$20 incentive, or no incentive. The results of the experiment were assessed by examining descriptive tables and analyzing logistic regression models.

**RESULTS/CONCLUSIONS:** Overall, respondents in the incentive group had higher cooperation rates while lowered data collection costs. The increased response rate did not significantly change the population estimates for drug abuse. The results of the logit model revealed different levels of cooperation for different demographic subgroups. However, the incentives neither enhanced nor reduced the difference in levels of cooperation across subgroups. The results indicate that it was beneficial for the survey to use incentives to encourage cooperation.

### **Processing of race and ethnicity in the 2004 National Survey on Drug Use and Health**

**CITATION:** Grau, E. A., Martin, P., Frechtel, P., Snodgrass, J., & Caspar, R. (2005). Processing of race and ethnicity in the 2004 National Survey on Drug Use and Health. In *Proceedings of the*

*2005 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Minneapolis, MN (pp. 3076-3083). Alexandria, VA: American Statistical Association.*

**PURPOSE/OVERVIEW:** Since the inception of NSDUH, the race and ethnicity of each respondent have been included. They are used as part of the demographic breakdowns in the analyses and the various reports generated from the survey. From 1971 to 1998, the race and ethnicity questions underwent few changes. However, along with the switch from paper-and-pencil interviewing (PAPI) methods of questionnaire administration to computer-assisted interviewing (CAI) methods in 1999, the race and ethnicity categories were updated pursuant to new Office of Management and Budget (OMB) directives. This paper details how race and ethnicity data were recorded in NSDUH since the 1999 CAI and summarizes how these data were processed.

**METHODS:** N/A.

**RESULTS/CONCLUSIONS:** N/A.

## **Development of a Spanish questionnaire for NSDUH**

**CITATION:** Hinsdale, M., Díaz, A., Salinas, C., Snodgrass, J., & Kennet, J. (2005). Development of a Spanish questionnaire for NSDUH. In J. Kennet & J. Gfroerer (Eds.), *Evaluating and improving methods used in the National Survey on Drug Use and Health* (HHS Publication No. SMA 05-4044, Methodology Series M-5, pp. 89-104). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** Translation of survey questionnaires is becoming standard practice for large-scale data collection efforts as growing numbers of immigrants arrive in the United States. However, the methods used to produce survey translations have not been standardized—even for Spanish, the most common target language (Shin & Bruno, 2003). The translation review of the National Survey on Drug Use and Health (NSDUH) was carried out for a variety of reasons. For many years, the survey has provided a Spanish-language version of the instrument for respondents who requested it. Each year, as new questions were added to the survey, translations were carried out on an ad hoc basis using a variety of translators. In the 1999 survey redesign, a large number of questions were added, and a large number of existing questions were altered to accommodate the audio computer-assisted self-interviewing (ACASI) format. It became apparent through feedback from the field that some of the Spanish questions seemed awkward; consequently, survey staff decided that a comprehensive review would be appropriate. It was determined that a multicultural review of the 2000 survey's Spanish instrument would be the most effective procedure.

**METHODS:** This chapter describes the techniques and principles that were applied in a multicultural review of the translation of the NSDUH questionnaire. Common problems that arose in the translation process and "best practices" for their resolution also are discussed. In addition, because increasing numbers of surveys are employing computer-assisted interviewing (CAI), this chapter illustrates some of the ways in which this technology ideally can be put to use in conveying translated materials. Using three translators coming from varied

backgrounds in Central and South America, and focus groups of potential respondents representing the major Spanish-speaking groups in the United States, a translation service that specialized in this type of work carried out a review of the entire questionnaire. The specifics of the focus group and multicultural review processes that took place are described in this chapter within the context of a discussion of best practices for the development of Spanish survey translations.

**RESULTS/CONCLUSIONS:** Several critical steps in the development of accurate Spanish survey translations were identified by the authors. A seemingly obvious first step involves staffing the project with qualified personnel—from translators to interviewers. To optimize respondent comprehension, translations should be developed using a multicultural approach and should be tested and reviewed thoroughly by a diverse group of bilingual individuals (Schoua-Glusberg, 2000). Understanding and applying the concept of cognitive equivalence versus direct translation are key in the development of an effective survey translation. Just as in questionnaire development of English-language surveys, cognitive testing should be employed to identify and correct potential flaws in wording. For studies such as NSDUH that use ACASI, a professional Spanish-speaking voice and skilled audio technicians are needed to ensure the high quality of the audio recording, which maximizes respondents' comprehension. Bilingual interviewers should be fluent and literate in both Spanish and English, and these skills must be demonstrated using a standardized certification procedure. Finally, allowing sufficient time to implement the Spanish translation and train the interviewing staff is perhaps the most problematic step of all because data collection schedules are typically rigorous and researchers are often challenged to maintain the critical timeline even without translations.

## **Results of the variance component analysis of sample allocation by age in the National Survey on Drug Use and Health**

**CITATION:** Hunter, S. R., Bowman, K. R., & Chromy, J. R. (2005). Results of the variance component analysis of sample allocation by age in the National Survey on Drug Use and Health. In *Proceedings of the 2005 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Minneapolis, MN* (pp. 3132-3136). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** Since 1999, person sampling rates for the National Survey on Drug Use and Health (NSDUH) have been set by State for five age groups: 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older. The sample design requires equal sample sizes of 22,500 individuals for each of three age groups: 12 to 17, 18 to 25, and 26 or older. The sample allocation of 22,500 adults to the three 26 or older age groups was set in 1999, then adjusted for the 2001 sample. Using parametric variance modeling, the sampling statistician can represent the variance of key estimates as a function of sample design parameters. This paper examines some alternative sample allocations to age groups based on an update of variance model parameters for nine key NSDUH estimates.

**METHODS:** Data from the 2003 NSDUH were used to estimate the parameters needed for these models. Because many aspects of the sample design were fixed by the 5-year coordinated design (e.g., number of sampling units and total sample size by State), the models focused on two goals:

(1) to predict the expected variance for the 2005 study for selected measures, and (2) to review the allocation of the 26 or older sample among those aged 26 to 35, 36 to 49, and 50 or older.

**RESULTS/CONCLUSIONS:** The results showed a larger percentage of the population in the older age groups and increasing use and dependence in these age groups.

### **Forward telescoping bias in reported age of onset: An example from cigarette smoking**

**CITATION:** Johnson, E. O., & Schultz, L. (2005). Forward telescoping bias in reported age of onset: An example from cigarette smoking. *International Journal of Methods in Psychiatric Research*, 14(3), 119-129.

**PURPOSE/OVERVIEW:** Age at the onset of a disorder is a critical characteristic that may predict the increased risk of a severe course and genetic liability. However, retrospectively reported onset in surveys is subject to measurement error. This article investigates forward telescoping, a bias in which respondents report events closer to the time of interview than is true. Past research suggests that forward telescoping influences reported age at onset of first substance use, but it does not answer other questions, such as "Is there a difference in the influence of forward telescoping on age of initiation between experimental users (those who have ever used drugs or smoke on a regular basis) and those who use drugs regularly?" or "Does forward telescoping affect reported age at onset for more advanced stages of substance use?" Thus, the purpose of this paper is to examine the effect of this bias on age of onset for smoking initiation and daily smoking.

**METHODS:** To estimate the effect of age at interview independent of birth year cohort based on multiple cross-sectional surveys of the same population, the authors selected respondents born between 1966 and 1977 ( $n = 82,122$ ) from the 1997–2000 National Household Surveys on Drug Abuse (NHSDA). Logistic regression was used to estimate the magnitude of forward telescoping in reported age when the first cigarette was smoked to test whether forward telescoping was greater for experimental smokers than for regular smokers and to assess whether the magnitude of forward telescoping in reported age of first daily smoking was lower than that of initiation.

**RESULTS/CONCLUSIONS:** The authors found an association between age at onset and age at interview, within birth year, for experimenters and for daily smokers. In addition, as age at interview increased from 12 to 25, the authors found that the probability of reporting early onset decreased by half. Contrary to the hypothesis, forward telescoping of age at initiation appeared to affect equally both experimental smokers and daily smokers. However, it also was found that the degree of forward telescoping of age at initiation of smoking differed significantly by gender and that significant forward telescoping of age at onset of daily smoking occurred differentially by race/ethnicity. Overall, the results of the analysis suggest that forward telescoping is a nonignorable bias that can possibly be mitigated through attention to components of the survey design process, such as question and sample design.

## **Evaluating and improving methods used in the National Survey on Drug Use and Health**

CITATION: Kennet, J., & Gfroerer, J. (Eds.). (2005). *Evaluating and improving methods used in the National Survey on Drug Use and Health* (HHS Publication No. SMA 05-4044, Methodology Series M-5). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

PURPOSE/OVERVIEW: The National Survey on Drug Use and Health (NSDUH) is the leading source of information on the prevalence and incidence of the use of alcohol, tobacco, and illicit drugs in the United States. It is currently administered to approximately 67,500 individuals annually, selected from the civilian, noninstitutionalized population of the United States, including Alaska and Hawaii. A survey of this size and importance is compelled to utilize the latest and best methodology over all facets of its operations. Given the sample size and the careful sampling procedures employed, NSDUH provides fertile soil for the testing and evaluation of new methodologies. Evaluation of NSDUH methodologies has been and continues to be an integral component of the project. This includes not only reviewing survey research literature and consulting with leading experts in the field, but also conducting specific methodological studies tailored to the particular issues and problems faced by this survey (Gfroerer, Eyerman, & Chromy, 2002; Turner, Lessler, & Gfroerer, 1992).

METHODS: This volume provides an assortment of chapters covering some of the recent methodological research and development in NSDUH, changes in data collection methods and instrument design, as well as advances in analytic techniques. As such, it is intended for readers interested in particular aspects of survey methodology and is a must-read for those with interests in analyzing data collected in recent years by NSDUH.

RESULTS/CONCLUSIONS: This volume contains a collection of some recent methodological research carried out under the auspices of the NSDUH project. Publishing these studies periodically provides a resource for survey researchers wishing to catch up on the latest developments from this unique survey. Readers from a variety of backgrounds and perspectives will find these chapters interesting and informative and, it is hoped, useful in their own careers in survey methodological research, drug abuse prevention and treatment, and other areas.

### **Introduction. In J. Kennet & J. Gfroerer (Eds.), Evaluating and improving methods used in the National Survey on Drug Use and Health**

CITATION: Kennet, J., & Gfroerer, J. (2005). Introduction. In J. Kennet & J. Gfroerer (Eds.), *Evaluating and improving methods used in the National Survey on Drug Use and Health* (HHS Publication No. SMA 05-4044, Methodology Series M-5, pp. 1-6). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

PURPOSE/OVERVIEW: The National Survey on Drug Use and Health (NSDUH) is the leading source of information on the prevalence and incidence of the use of alcohol, tobacco, and illicit drugs in the United States. It is currently administered to approximately 67,500 individuals annually, selected from the civilian, noninstitutionalized population of the United States,

including Alaska and Hawaii. A survey of this size and importance is compelled to utilize the latest and best methodology over all facets of its operations. Given the sample size and the careful sampling procedures employed, NSDUH provides fertile soil for the testing and evaluation of new methodologies. Evaluation of NSDUH methodologies has been and continues to be an integral component of the project. This includes not only reviewing survey research literature and consulting with leading experts in the field, but also conducting specific methodological studies tailored to the particular issues and problems faced by this survey (Gfroerer, Eyerman, & Chromy, 2002; Turner, Lessler, & Gfroerer, 1992). This volume provides an assortment of chapters covering some of the recent methodological research and development in NSDUH, changes in data collection methods and instrument design, as well as advances in analytic techniques.

**METHODS:** This introduction begins with a brief history and description of NSDUH. A more detailed account can be found in Gfroerer et al. (2002). Prior methodological research on NSDUH then is described, followed by an account of the major methodological developments that were implemented in 2002. Finally, each of the chapters and their authors are introduced.

**RESULTS/CONCLUSIONS:** This volume contains a collection of some recent methodological research carried out under the auspices of the NSDUH project. Publishing these studies periodically provides a resource for survey researchers wishing to catch up on the latest developments from this unique survey. Readers from a variety of backgrounds and perspectives will find these chapters interesting and informative and, it is hoped, useful in their own careers in survey methodological research, drug abuse prevention and treatment, and other areas.

## **Introduction of an incentive and its effects on response rates and costs in NSDUH**

**CITATION:** Kennet, J., Gfroerer, J., Bowman, K. R., Martin, P. C., & Cunningham, D. (2005). Introduction of an incentive and its effects on response rates and costs in NSDUH. In J. Kennet & J. Gfroerer (Eds.), *Evaluating and improving methods used in the National Survey on Drug Use and Health* (HHS Publication No. SMA 05-4044, Methodology Series M-5, pp. 7-18). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** In 2002, the National Survey on Drug Use and Health (NSDUH) began offering respondents a \$30 cash incentive for completing the questionnaire. This development occurred within the context of a name change and other methodological improvements to the survey and resulted in significantly higher response rates. Moreover, the increased response rates were achieved in conjunction with a net decrease in costs incurred per completed interview. This chapter presents an analysis of response rate patterns by geographic and demographic characteristics, as well as interviewer characteristics, before and after the introduction of the incentive. Potential implications for other large-scale surveys also are discussed.

**METHODS:** To demonstrate the effects of the incentive and other changes, a comparison is presented of the response rates, by quarters, in the years before and after the incentive was

introduced. These rates then are broken down by geographic area, population density, respondent age, gender, and race/ethnicity. Response rates also are examined with respect to the characteristics of interviewers, including their prior experience on the survey, race/ethnicity, and gender.

**RESULTS/CONCLUSIONS:** The \$30 incentive, with possible help from the other changes that were introduced in January 2002, produced dramatic improvement in the number of eligible respondents who agreed to complete the NSDUH interview. Moreover, the increase in respondent cooperation was accompanied by a decrease in cost per completed interview. Clearly, the adoption of the incentive was beneficial to all involved, with the possible exception of the field interviewers, who required fewer hours to complete their assignments on the project and consequently received less pay.

From these analyses, it appears that incentives had their most pronounced effect among people between the ages of 12 and 25. Because these are known to be the years in which substance use and abuse are most prevalent and have their greatest incidence, it seems the incentive was well spent in terms of capturing the population of most interest. However, serious thought needs to be given to methods for attracting those older than 25. It could be the case that \$30 was simply insufficient to attract people who have settled into careers and/or other more rewarding activities, such as child rearing or retirement, or it could be that these people did not participate for other reasons. These reasons will have to be investigated and addressed in order for NSDUH to optimally cover the aging baby-boom generation and other cohorts.

### **Applying cognitive psychological principles to the improvement of survey data: A case study from the National Survey on Drug Use and Health**

**CITATION:** Kennet, J., Painter, D., Barker, P., Aldworth, J., & Vorburger, M. (2005). Applying cognitive psychological principles to the improvement of survey data: A case study from the National Survey on Drug Use and Health. In *Proceedings of the 2005 Joint Statistical Meetings, American Statistical Association, AAPOR - Section on Survey Research Methods, Minneapolis, MN* (pp. 3887-3897). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** The National Survey on Drug Use and Health (NSDUH) collects data on Medicare and Medicaid coverage as part of a general interview conducted after the core drug use measures have been administered. Although the overall estimates derived from the NSDUH Medicare and Medicaid coverage questions have generally appeared credible, it became apparent that among individuals younger than 65 years old, Medicare coverage was overreported and Medicaid coverage was underreported. Among adults aged 65 or older, Medicaid coverage appeared to be highly overreported. These judgments were based on "eyeball" comparisons with estimates from the Survey of Income and Program Participation (SIPP) and the Current Population Survey (CPS), both of which administered more detailed modules on health insurance coverage.

**METHODS:** The Medicare and Medicaid questions were subject to expert reviews within the context of Tourangeau, Rips, and Rasinski's (2000) Response Process Model, which posits four processes involved in answering survey questions: comprehension, retrieval, judgment, and

response. Three reviewers independently critiqued the questions. Two were survey methodologists having extensive experience with NSDUH's content and fielding. The third reviewer was a cognitive psychologist who was new to the project. The review team met several times over the course of a few days to compare comments and draft a revised pair of items.

**RESULTS/CONCLUSIONS:** Expert review of the NSDUH question wordings suggested that inadequate establishment of context (i.e., defining terms after using them in the questions) and other syntactic difficulties created excessive demands on working memory. Correction of these problems in the 2003 NSDUH resulted in age group coverage estimates that more closely matched those obtained in the other surveys, which targeted this topic more specifically.

### **Assessing the reliability of key measures in the National Survey on Drug Use and Health using a test-retest methodology**

**CITATION:** Kennet, J., Painter, D., Hunter, S. R., Granger, R. A., & Bowman, K. R. (2005, November). *Assessing the reliability of key measures in the National Survey on Drug Use and Health using a test-retest methodology*. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC.

**PURPOSE/OVERVIEW:** The National Survey on Drug Use and Health (NSDUH) is a major source of information on substance use and mental illness prevalence in the United States. It is administered in households to approximately 67,500 individuals annually using a complex, multistage sampling design. Assessing the reliability of estimates produced by NSDUH is of primary importance to those who use these data for research and in the making of policy decisions. In quarters 1 and 2 of 2005, a pretest was carried out in which approximately 200 NSDUH respondents were reinterviewed as an effort to refine the methods to be used in conducting a large-scale reliability field test in 2006. This paper discusses the design and procedural considerations that were taken into account in planning the pretest and upcoming field test.

**METHODS:** Considerations included the time interval between the test and pretest, the sample size needed for reliability estimates of low prevalence behaviors, whether the sample would be embedded or not in the NSDUH main study, using the same versus different interviewers for the reinterview, increased risk of loss of respondent privacy due to the provision of recontact information, amount of incentive for the reinterview, and others. In addition, preliminary findings from the pretest were discussed that may influence methods employed in the 2006 field test, such as response rates on the reinterview and respondent feedback.

**RESULTS/CONCLUSIONS:** The reliability study pretest achieved higher than expected reinterview response rates, successfully completed reinterviews within the 5- to 15-day window, displayed a high level of consistency in responses to drug and demographic questions between T1 and T2, received a positive response from respondents, and demonstrated that field interviewers will be able to follow the procedures and protocols in the 2006 reliability study.

## **The use of monetary incentives in federal surveys on substance use and abuse**

CITATION: Kulka, R. A., Eyerman, J., & McNeeley, M. E. (2005). The use of monetary incentives in federal surveys on substance use and abuse. *Journal for Social and Economic Measurement*, 30(2-3), 233-249.

PURPOSE/OVERVIEW: Empirical research over the past 30 years has shown positive effects of monetary incentives on response rates. This paper discusses the use of incentives specifically for surveys on substance use and abuse.

METHODS: This paper starts by providing a background and review of the current empirical literature on the effects that monetary incentives have on survey response rates, survey statistics, and other practical and operational issues in surveys. Next, two controlled experiments on the effect of incentives on substance use surveys are discussed: the Alcohol and Drug Services Study (ADSS) and the National Household Survey on Drug Abuse (NHSDA). Each of the studies randomized respondents into different incentive categories, including no incentives and two to three levels of increasing incentives.

RESULTS/CONCLUSIONS: The ADSS results revealed that higher incentives correlated with higher cooperation rates, but that the difference in cooperation rates between the levels of incentives was small. The analyses also revealed that the incentives did not affect different subgroups differently. In addition, the results indicated that the use of incentives did not affect the quality of respondents' answers. The NHSDA results, however, showed that although incentives did increase response rates, they had differing impacts on different subgroups. The results of the analyses on survey response bias were inconclusive. The results of these two studies revealed that more research is needed on this topic to further understand the effect of incentives on survey data quality. The results of these studies are described in more detail in a series of papers appearing in the *Journal of Social and Economic Measurement*.

## **Effects of the September 11, 2001, terrorist attacks on NSDUH response rates**

CITATION: McNeeley, M. E., Odom, D., Stivers, J., Frechtel, P., Langer, M., Brantley, J., Painter, D., & Gfroerer, J. (2005). Effects of the September 11, 2001, terrorist attacks on NSDUH response rates. In J. Kennet & J. Gfroerer (Eds.), *Evaluating and improving methods used in the National Survey on Drug Use and Health* (HHS Publication No. SMA 05-4044, Methodology Series M-5, pp. 31-58). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

PURPOSE/OVERVIEW: The purpose of this study was to determine whether the events of September 11th had an effect on screening response rates (SRRs) and interview response rates (IRRs).

METHODS: In addition to the preliminary statistical analysis of response rate data, logistic regression models were run to provide a more refined analysis. Field interviewer (FI) focus groups also were used to assess changes in the logistics of FI activity and in the use of the lead letter. To capture heightened concerns about security, FIs also were asked about increases in

controlled access problems and changes in the mode of contact with screening and interview respondents.

**RESULTS/CONCLUSIONS:** It was found that the New York City consolidated metropolitan statistical area (CMSA) and Washington, DC, primary metropolitan statistical area (PMSA) response rates suffered dramatic decreases following the September 11th terrorist attacks, although the differences in IRR in the Washington, DC, PMSA were shown to be significant only after modeling on a number of factors. The national screening response rate (SRR) also showed a decrease even after removing the New York City CMSA and Washington, DC, PMSA from the sample. This decrease was significant but less dramatic than in the two metropolitan areas.

### **Interviewer falsification detection using data mining**

**CITATION:** Murphy, J., Eyerman, J., McCue, C., Hottinger, C., & Kennet, J. (2005, October). *Interviewer falsification detection using data mining*. Presented at Statistics Canada's 22nd International Symposium Series, Ottawa, Ontario, Canada.

**PURPOSE/OVERVIEW:** Interviewer falsification is the deliberate creation of survey responses by the interviewer without input from the respondent. Undetected falsification can introduce bias into the population estimates if falsified responses do not match the values that would have been provided by respondents. Currently, the procedures required to detect interviewer fraud can be expensive and draw resources away from the study that could be applied to data quality procedures. Data mining can be used to program falsification propensity checks that may be run on a frequent basis, facilitating timely and relatively inexpensive detection and remediation, and serving as a possible deterrent to falsification.

**METHODS:** This paper describes an innovative use of data mining on response data and metadata to identify, characterize, and prevent falsification by field interviewers on the National Survey on Drug Use and Health (NSDUH).

**RESULTS/CONCLUSIONS:** This study yielded mixed results. It clearly demonstrated that there is potential for data mining to be used as a falsification detection tool on NSDUH. In particular, the complementary findings noted in different data resources underscored the value that can be obtained by using a combination of automated search strategies and expert review on the various NSDUH databases. For example, a putative association between difficult-to-complete interviews and falsification was supported by the record of calls (ROC) data, which suggested that the decision to falsify was made after an interviewer had experienced difficulty in contacting a subject, and the finding that an increased number of breakoffs was associated with interviewers who had been placed under increased scrutiny by the data quality team.

## **Appendix C: Research on the impact of changes in NSDUH methods**

**CITATION:** Office of Applied Studies. (2005). Appendix C: Research on the impact of changes in NSDUH methods. In *Results from the 2004 National Survey on Drug Use and Health: National findings* (HHS Publication No. SMA 05-4062, NSDUH Series H-28, pp. 145-154). Rockville, MD: Substance Abuse and Mental Health Services Administration.

**PURPOSE/OVERVIEW:** Although the design of the 2002 through 2004 National Surveys on Drug Use and Health (NSDUH) was similar to the design of the 1999 through 2002 surveys, there were important methodological differences between the 2002 to 2004 NSDUHs and prior surveys, including a change in the survey's name, the introduction of an incentive, improved data collection quality control procedures and the use of the 2000 decennial census for sample weighting. The results of the 2002 survey suggested that the incentive had an impact on estimates. A panel of survey methodology experts concluded that it would not be possible to measure the effects of each changes separately because of the multiple changes made to the survey simultaneously and recommended that the Substance Abuse and Mental Health Services Administration (SAMHSA) continue its analyses of the 2001 and 2002 data to learn as much as possible about the impacts of each of the methodological improvements. The purpose of this appendix is to summarize the studies of the effects of the 2002 method changes and to discuss the implications of this body of research for analysis of NSDUH data.

**METHODS:** Early analyses were presented to a panel of survey design and survey methodology experts convened on September 12, 2002. The analyses included (1) retrospective cohort analyses; (2) response rate pattern analyses; (3) response rate impact analyses; (4) analyses of the impact of new census data; and (5) model-based analyses of protocol changes, name change, and incentives. Since 2002, two additional analyses were conducted that extend those described above: more in-depth incentive experiment analyses and further analysis of the 2001 field interventions.

**RESULTS/CONCLUSIONS:** A summary of all of the results of the 2002 NSDUH method analyses was presented to a second panel of consultants on April 28, 2005. The panel concluded that there was no possibility of developing a valid direct adjustment method for the NSDUH data and that SAMHSA should not compare 2002 and later estimates with 2001 and earlier estimates for trend assessment. The panel suggested that SAMHSA make this recommendation to users of NSDUH data.

### **Analysis of NSDUH record of call data to study the effects of a respondent incentive payment**

**CITATION:** Painter, D., Wright, D., Chromy, J. R., Meyer, M., Granger, R. A., & Clarke, A. (2005). Analysis of NSDUH record of call data to study the effects of a respondent incentive payment. In J. Kennet & J. Gfroerer (Eds.), *Evaluating and improving methods used in the National Survey on Drug Use and Health* (HHS Publication No. SMA 05-4044, Methodology Series M-5, pp. 19-30). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** The National Survey on Drug Use and Health (NSDUH) employs a multistage area probability sample to produce population estimates of the prevalence of substance use and other health-related issues. Letters are sent to selected households to alert potential respondents of the interviewer's future visit. Interviewers then visit the residence to conduct a brief screening, which determines whether none, one, or two individuals are selected from the household. To maximize response rates, interviewers may make several visits to a household to obtain cooperation (the term "call" is used for "visits" from this point on with the understanding that this is a face-to-face survey; telephones are not used to contact potential respondents). In-person interviews are conducted with selected respondents using both computer-assisted personal interviewing (CAPI) and audio computer-assisted self-interviewing (ACASI). Sensitive questions, such as those on illicit drug use, are asked using the ACASI method to encourage honest reporting. A detailed description of the NSDUH methodology is described elsewhere (RTI International, 2003). During the late 1990s, NSDUH experienced a slight decline in response rates. A closer examination of the data revealed stable noncontact patterns, but increasing refusal rates (Office of Applied Studies [OAS], 2001). This implied that sample members were beginning to become less likely to participate once they were contacted. This was compounded by the need to hire a large number of new interviewers who may not have had the confidence or skills to overcome respondent refusals.

**METHODS:** Given the slight decline in response rates, and the expectation that this trend might become more serious, NSDUH staff designed an experiment to evaluate the effectiveness of monetary incentives in improving respondent cooperation. A randomized, split-sample, experiment was conducted during the first 6 months of data collection in 2001. The experiment was designed to compare the impact of \$20 and \$40 incentive treatments with a \$0 control group on measures of respondent cooperation and survey costs.

**RESULTS/CONCLUSIONS:** The results showed that both the \$20 and \$40 incentives increased overall response rates while producing significant cost savings when compared with the \$0 control group (Eyerman, Bowman, Butler, & Wright, 2002). Preliminary analysis showed no statistically detectable effects of the three incentive treatments on selected substance use estimates. Subsequent analysis showed some positive and negative effects depending on the substance use measure when the \$20 and \$40 treatments were combined and compared with the \$0 control group (Wright, Bowman, Butler, & Eyerman, 2002). Based on the outcome of the 2001 experiment, NSDUH staff implemented a \$30 incentive in 2002. Their analysis showed that a \$30 incentive would strike a balance between gains in response rates and cost savings. This chapter analyzes the effect of the new \$30 incentive on the data collection process as measured by record of calls (ROC) information. The effect of the incentives implemented in 2002 on response rates and costs is discussed by Kennet et al. in Chapter 2 of this volume.

## **Analyzing audit trails in NSDUH**

**CITATION:** Penne, M. A., Snodgrass, J., & Barker, P. (2005). Analyzing audit trails in NSDUH. In J. Kennet & J. Gfroerer (Eds.), *Evaluating and improving methods used in the National Survey on Drug Use and Health* (HHS Publication No. SMA 05-4044, Methodology Series M-5, pp. 105-120). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** For the National Survey on Drug Use and Health (NSDUH), the interview sections on substance use and other sensitive topics were changed in 1999 from self-administered paper-and-pencil interviewing (PAPI) to audio computer-assisted self-interviewing (ACASI). These changes were prompted by research that showed that computer-assisted interviewing (CAI) questionnaires reduced input errors. Research also showed that use of ACASI increased comprehension for less literate respondents and, by increasing privacy, resulted in more honest reporting of illicit drug use and other sensitive behaviors.

**METHOD:** In this chapter, the earlier work is briefly described, possible methods for streamlining the data-processing portion are discussed, and the use of audit trails in the 2002 survey is focused on to investigate three aspects of data quality: question timing, respondent breakoffs, and respondent "backing up" to change prior responses.

**RESULTS/CONCLUSIONS:** Timing data showed that when measured against a gold standard (GS) time, field interviewers (FIs) were spending approximately the correct amount of time with the very beginning of the interview at the introduction to the CAI instrument screen. However, once past this point, they spent less time than the GS on several important aspects of the questionnaire, such as setting up the calendar, setting up the ACASI tutorial, completing the verification form, and ending the interview with the respondent. Conversely, they were taking longer than the GS in ending the ACASI portion of the interview.

## **Modeling context effects in the National Survey of Drug Use and Health (NSDUH)**

**CITATION:** Wang, K., Baxter, R., & Painter, D. (2005). Modeling context effects in the National Survey of Drug Use and Health (NSDUH). In *Proceedings of the 2005 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Minneapolis, MN* (pp. 3646-3651). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** Context effects occur for multiple reasons. Primarily, the response to a question is affected by information that is not part of the question itself, and the cognitive process has been affected because of the content of the preceding questions. In terms of questionnaire changes, context effects may be said to take place between two survey questions when a change introduced to the first (or contextual) item affects the response process for the subsequent (target) item, which in turn may lead to a different response than if the change had not been made. Comparatively little work has been done to examine if different types of respondents might be more or less susceptible to changes in context.

**METHODS:** In this paper, the authors used data from the National Survey on Drug Use and Health (NSDUH) for 2002 and 2003 to determine if some types of respondents were more greatly affected by a contextual change than other respondents.

**RESULTS/CONCLUSIONS:** The authors found that the removal of item SEN13A in the 2003 NSDUH had an effect on responses to item SEN13B in 2003 as compared with previous years. What remained unclear was the nature of the means by which the removal of SEN13A affected responses to SEN13B. The estimated models identified current cigarette users (who were not

also current marijuana users) as respondents who were especially likely to respond to item SEN13B with responses of "neither approve nor disapprove" in 2003 than in 2002. The models also correctly identified respondents who previously used cigarettes and had never used marijuana as more likely to respond to SEN13B with "neither approve nor disapprove" in 2003 than in 2002.

### **Are two feet in the door better than one? Using process data to examine interviewer effort and nonresponse bias**

CITATION: Wang, K., Murphy, J., Baxter, R., & Aldworth, J. (2005, November). *Are two feet in the door better than one? Using process data to examine interviewer effort and nonresponse bias*. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC.

PURPOSE/OVERVIEW: The authors examined the use of administrative call record data from the National Survey on Drug Use and Health (NSDUH) in order to address interviewing issues. The authors first described NSDUH and the available call record data, then examined how the calling strategies of interviewers can affect contact and cooperation rates. The authors also conducted analyses to examine the relationships between the volume of call attempts and survey estimates and, in turn, the potential for bias due to nonresponse.

METHODS: The authors took the first steps in analyzing NSDUH process data to address questions regarding interviewer efforts and effects on response rates and survey estimates.

RESULTS/CONCLUSIONS: The authors found that calling times, defined by the time of the call (before or after 4:00 p.m.) and the day of the week (weekday vs. weekend) were related to contact on the first attempt for the screener. They also found evidence that using less intensive follow-up efforts would not necessarily lead to survey estimates that differed significantly from estimates obtained with greater effort. They also suggested that reduction of the interviewing effort on a per case basis could lead to reductions in data collection costs.

### **Decomposing the total variation in a nested random effects model of neighborhood, household, and individual components when the dependent variable is dichotomous: Implications for adolescent marijuana use**

CITATION: Wright, D., Bobashev, G. V., & Novak, S. P. (2005). Decomposing the total variation in a nested random effects model of neighborhood, household, and individual components when the dependent variable is dichotomous: Implications for adolescent marijuana use. *Drug and Alcohol Dependence*, 78(2), 195-204.

PURPOSE/OVERVIEW: Multilevel modeling techniques have become a useful tool that enables substance abuse researchers to more accurately identify the contribution of multiple levels of influence on drug-related attitudes and behaviors. However, it is difficult to determine the relative importance of the different hierarchical levels because the variance components

estimation involves calculations using a log-odds metric at the lowest level of estimation in the case of dichotomous outcomes,

**METHODS:** The authors presented methods that were introduced by Goldstein and Rasbash (1996) to convert the variance components from the log-odds to the probability metric. These methods have a few advantages in that they provide a more logical and interpretable way to examine variation for nonlinear outcomes, which tend to be heavily utilized in substance use research. With data from the 1999 National Household Survey on Drug Abuse, the authors partitioned variation among individual, household, and neighborhood levels for the binary outcome of past year marijuana use to illustrate this approach. The authors also conducted a stability analysis to examine the robustness across different estimation procedures commonly available in commercial multilevel software packages. Furthermore, the authors partitioned the variance components using a conventional continuously distributed outcome and compared the relative magnitudes across binary and continuous outcomes.

**RESULTS/CONCLUSIONS:** The authors reported that both binary and continuous indicators of drug use could be used for characterizing use within households and neighborhoods in a similar statistical way providing interpretable results. They demonstrated that the inverse logit method was applicable to any number of hierarchical levels and variable sizes of the clusters.

### **Non-response bias from the National Household Survey on Drug Abuse incentive experiment**

**CITATION:** Wright, D., Bowman, K., Butler, D., & Eyerman, J. (2005). Non-response bias from the National Household Survey on Drug Abuse incentive experiment. *Journal for Social and Economic Measurement*, 30(2-3), 219-231.

**PURPOSE/OVERVIEW:** In a preliminary experiment conducted in the 2001 National Household Survey on Drug Abuse (NHSDA), it was concluded that providing incentives increased response rates; therefore, a \$30 incentive was used in the subsequent 2002 NHSDA. The purpose of this paper is to explore the effect that incentive had on nonresponse bias.

**METHODS:** The sample data were weighted by likelihood of response between the incentive and nonincentive cases. Next, a logistic regression model was fit using substance use variables and controlling for other demographic variables associated with either response propensity or drug use.

**RESULTS/CONCLUSIONS:** The results indicate that for past year marijuana use, the incentive is either encouraging users to respond who otherwise would not respond, or it is encouraging respondents who would have participated without the incentive to report more honestly about drug use. Therefore, it is difficult to determine whether the incentive money is reducing nonresponse bias, response bias, or both. However, reports of past year and lifetime cocaine did not increase in the incentive category, and past month use of cocaine actually was lower in the incentive group than in the control group.

## 2006

### **Variance estimation of the survey-weighted kappa measure of agreement**

CITATION: Feder, M. (2006). Variance estimation of the survey-weighted kappa measure of agreement. In *Proceedings of the 2006 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Seattle, WA* [CD-ROM] (pp. 3002-3007). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: A National Survey on Drug Use and Health (NSDUH) interview/reinterview study was conducted to assess the reliability of the responses (T1 and T2) by comparing them. To measure the reliability of categorical responses, Cohen's kappa index of interrater reliability, which is the most often used statistic to assess interrater reliability of categorical variables, was used. The common variance estimation approach is to use the Fleiss et al. (1969) asymptotic variance formula, which assumes an independent sample, with equal probabilities of inclusion. However, NSDUH's sample design is complex, involving clustering and unequal weighting to account for variable probabilities of inclusion and nonresponse adjustments. Therefore, this may have a significant effect on the point estimates of kappa and the estimation of its variance. Although correcting the point estimates of kappa for the design was straightforward, the variance estimate was more involved.

METHODS: The author presented a Taylor linearization (TL) derivation, along with simulation results of the TL method and of the Fleiss et al. (1969) formula, and their assessment.

RESULTS/CONCLUSIONS: Results from this small simulation study demonstrated the bias of the standard formula when clustering was present, and the good performance of the proposed procedure.

### **Population coverage in the National Survey on Drug Use and Health**

CITATION: Morton, K. B., Hunter, S. R., Chromy, J. R., & Martin, P. C. (2006). Population coverage in the National Survey on Drug Use and Health. In *Proceedings of the 2006 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Seattle, WA* (pp. 3441-3446). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: The National Survey on Drug Use and Health (NSDUH) examined coverage and response rates for various populations using data from the 1993 through 1998 National Household Surveys on Drug Abuse (NHSDAs). Several design and methodological changes have been implemented since then. The motivation for this paper was to see if the design and methodological changes had any impact on coverage and response rates and to update the prior analysis using data from the 1999 through 2001 NHSDAs and the 2002 through 2004 NSDUHs.

METHODS: The study examined coverage in a broad sense by looking at screening and interview response rates, coverage rates, and dwelling unit eligibility rates for completeness,

even though this is not part of coverage. The dwelling unit eligibility rate was defined as the total number of eligible units divided by the total number of selected units. The screening response rate was the number of completed screeners divided by the number of eligible households. At the person level, the interview response rate was computed as the number of respondents divided by the number of selected individuals. Finally, the coverage rate was computed as the ratio of the weighted survey estimates after adjustment for both screening and interview nonresponse (but before poststratification and extreme weight trimming) to the external and more precise estimates.

**RESULTS/CONCLUSIONS:** In general, NSDUH achieved very good screening and interview response rates and had good coverage of demographic subgroups. Since the 1993 through 1998 NHSDAs, the NSDUH experienced some variation in these rates because of the implementation of several design and methodological changes. Beginning in 1999, the 50-State design caused moderate decreases in overall dwelling unit eligibility, screening response, and interview response rates, while the coverage was about the same. Further, the methodological changes that were introduced in the 2002 survey caused increases in interview response and coverage rates for some domains. As seen in the consistently high coverage rates for Hispanics, NSDUH either overcounted this population, or the intercensal projections were low for this group. Finally, high coverage rates in the "Other" race category may have been due to the growing multiracial population and increased acceptability of specifying more than one race.

## **National Survey on Drug Use and Health: Field interviewer behavior and response rate analysis (FIBRRA) final report**

**CITATION:** Safir, A., Murphy, J., Wang, K., & Park, H. (2006, December). *National Survey on Drug Use and Health: Field interviewer behavior and response rate analysis (FIBRRA) final report* (prepared for the Substance Abuse and Mental Health Services Administration). Research Triangle Park, NC: RTI International.

**PURPOSE/OVERVIEW:** Changes in the composition of environmental, household or dwelling unit (DU), respondent, and field interviewer (FI) characteristics across survey phases affect response rates. This report explores the effects of DU, FI, calling pattern, respondent, and area-level factors on survey contact and cooperation in the National Survey on Drug Use and Health (NSDUH). The purpose of the research was to further the development of tools to use information generated during data collection (i.e., paradata and process data, such as contact attempt patterns) or in the advancement of data collection (i.e., auxiliary data, such as the characteristics of selected segments) to predict and improve response rates in upcoming quarters.

**METHODS:** The authors first modeled the relationship between screener contact, screener cooperation, and main interview cooperation, as well as multilevel factors such as county, segment, DU, respondent, and FI characteristics. These factors were selected in part based on the NSDUH sample design, which employs a multistage, stratified sample of primary sampling units, whether counties or groups of counties, segments (blocks or block groups), DUs, and individuals. In addition to contingency table and multiple logistic regression, multilevel modeling was used to account for the hierarchical nature of the data, including the effect of FI behavior and characteristics. To evaluate the practical application of these results, the authors

then examined factors identified by the multilevel modeling as having a statistically significant relationship with State-level screener interview and main interview cooperation rates across all States and quarters in 2004 in terms of their relationship to response rate variation in States that exhibited the largest change over 2004's quarter 3 to quarter 4 period. In reviewing the results, the authors assessed the effectiveness of the models in identifying factors associated with the outcome of the survey request and their usefulness in planning field strategies to address data collection challenges.

**RESULTS/CONCLUSIONS:** In the screener contact analysis, increases in population density, the percentage of the population aged 65 years or older, and the FI's experience were associated with higher contact rates. Single-unit DU building type and contact attempts on weekdays at 5:00 p.m. or later were also associated with higher contact rates. The screener contact was negatively associated with increases in the percentage of the segment who were aged 0 to 19 years old, the percentage of one-person households, the combined median/rent/home value, the FI's race (black or African American), and the presence of controlled access barriers.

In the screener cooperation analysis, increases in population density, median household income, and the presence of controlled access barriers were negatively associated with cooperation at first contact, controlling for other factors. In the screener cooperation analysis that was restricted to cases with pending refusals, increases in the driving under the influence (DUI) arrest rate and group quarters status were associated with higher cooperation. Increases in the percentage of DUs built between 1940 and 1949, median household income, FI age group, and total number of pending refusals were associated with lower cooperation.

In the main interview cooperation analysis, an increase in the percentage of the population aged 20 to 24 years old, FI race (black), respondent Hispanicity, and respondent gender (female) were associated with an increase in cooperation in the overall model. Conversely, an increase in the percentage of the population aged 35 to 44 years old, median household income, FI efficiency, FI hours per interview, total number of DUs assigned to an FI, and respondent age were associated with lower main interview cooperation. Further State-level analyses were unable to demonstrate the immediate utility of the factors identified in the multiple regression cooperation level analysis for making a univariate statement about the degree or direction of changes in response rates.

### **Screening for serious mental illness in populations with co-occurring substance use disorders: Performance of the K6 scale**

**CITATION:** Swartz, J. A., & Lurigio, A. J. (2006). Screening for serious mental illness in populations with co-occurring substance use disorders: Performance of the K6 scale. *Journal of Substance Abuse Treatment*, 31(3), 287-296.

**PURPOSE/OVERVIEW:** Serious mental illness (SMI), such as schizophrenia, bipolar disorder, and major depression, is prevalent among individuals with substance use disorders, particularly those in drug treatment programs. No screening tool has yet become the gold standard for identifying SMI among individuals with substance use disorders. One candidate instrument, the

K6 screening scale, is brief, easy to administer and score, and has performed well, detecting SMI in studies using general population samples.

**METHODS:** The authors used data from the National Survey on Drug Use and Health (NSDUH) to examine the K6's psychometric properties in a subsample of individuals with substance use disorders and found that the K6 accurately screened for severe psychological distress associated with SMI among individuals with substance use disorders and across different psychiatric disorders.

**RESULTS/CONCLUSIONS:** The findings suggested that the K6 is an accurate screening tool for assessing the likely presence of SMI among individuals with substance use disorders. The K6 performed well in detecting psychiatric disorders and was effective in identifying individuals with severe psychiatric impairment and a need for further assessment and intervention.

## 2007

### **Using callback models to adjust for nonignorable nonresponse in face-to-face surveys**

CITATION: Biemer, P., & Wang, K. (2007). Using callback models to adjust for nonignorable nonresponse in face-to-face surveys. In *Proceedings of the 2007 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Salt Lake City, UT* (pp. 2889-2896). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: In most sample surveys, weighting procedures attempt to compensate for nonresponse bias under the assumption of "ignorable nonresponse"; that is, the data that are known for both respondents and nonrespondents are sufficient to adequately adjust for nonresponse bias. For many surveys, call history data are available for all sample members, including nonrespondents, and because the level of effort (LOE) required to interview a sample member is likely to be highly correlated with response propensity, this method is ideally suited for modeling the nonignorable nonresponse. Biemer and Link (2007) provided a general method for nonresponse adjustment that relaxed the ignorable nonresponse assumption. Their method, which extended the ideas of Drew and Fuller (1980), used indicators of LOE based on call attempts to model the probability that an individual in the sample responds to the survey (referred to as the "response propensity"). Using data from the National Survey on Drug Use and Health (NSDUH), an annual in-person, cross-sectional study conducted in all 50 States and the District of Columbia to measure the prevalence and correlates of drug use in the U.S. population aged 12 or older, the authors investigated the feasibility to adjust for nonignorable nonresponse with callback models.

METHODS: The authors constructed several latent callback models consisting of four variables and compared them with an equivalent model that was constructed using the traditional logistic regression approach; that is, the model used was essentially the same the latent callback model with the special callback model features omitted.

RESULTS/CONCLUSIONS: Across a range of variables from NSDUH's screener, the callback model showed improvement over the traditional model as hypothesized. Although the magnitude of the improvements was not dramatic, the results clearly showed that gains in accuracy were possible using the special features of the callback model in response propensity weighting.

### **Patterns of nonresponse for key questions in NSDUH and implications for imputation**

CITATION: Frechtel, P., & Copello, E. (2007). Patterns of nonresponse for key questions in NSDUH and implications for imputation. In *Proceedings of the 2007 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Salt Lake City, UT* (pp. 3457-3464). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** The idea of using "soft nonrespondents" to represent "hard nonrespondents" is not new to survey research. Callbacks are often used to adjust for nonresponse in surveys. The goal is to control nonresponse bias by assuming that the hard nonrespondents are more similar to the callback respondents than they are to the original respondents.

**METHODS:** The National Survey on Drug Use and Health (NSDUH), an annual nationwide survey involving approximately 70,000 subjects per year, does not make use of callbacks. However, for several key questions in NSDUH, follow-up questions, or "probes," are presented to subjects who entered a response of "don't know" or "refused" to the original questions. The probes are intended to increase item response rates by simulating an actual interviewer. The probe respondents can be considered soft nonrespondents, and the subjects who answer neither the original question nor the probe can be viewed as hard nonrespondents. An analysis from an earlier study was expanded to include data pooled from the 2000 to 2005 surveys. The values of auxiliary variables were compared between the original respondents, the probe respondents, and the nonrespondents to see whether the nonrespondents resembled the probe respondents more than the original respondents.

**RESULTS/CONCLUSIONS:** The probes offered a less costly alternative to callbacks for the mitigation of nonresponse bias. The response patterns suggested that, for illicit drugs, subjects who refused to respond to the original question but responded to the probes were more often lifetime users (and more often recent users) than subjects who responded to the original question. The presence of the probes seemed to be correcting for some of the bias, simply by adjusting the estimates relative to what they would be if the probes did not exist. However, the probes could be further used in imputation, which would enhance the adjustment for the nonresponse bias. A comparison of the predicted means of the different response patterns suggested that, at least for lifetime marijuana use and lifetime cocaine use, the imputation method was able to pick up some, but not all, of the difference between the original respondents and the probe respondents.

### **Discrepancies in estimates of prevalence and correlates of substance use and disorders between two national surveys**

**CITATION:** Grucza, R. A., Abbacchi, A. M., Przybeck, T. R., & Gfroerer, J. C. (2007). Discrepancies in estimates of prevalence and correlates of substance use and disorders between two national surveys. *Addiction, 102*(4), 623-629.

**PURPOSE/OVERVIEW:** The purpose of this research was to assess the degree to which methodological differences might influence estimates of prevalence and correlates of substance use and disorders by comparing results from two surveys administered to nationally representative samples in the United States.

**METHODS:** The authors conducted a post hoc comparison of data from the 2002 National Survey on Drug Use and Health (NSDUH) with data from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) administered in 2001 and 2002.

**RESULTS/CONCLUSIONS:** Prevalence estimates for all substance use outcomes were higher in NSDUH than in NESARC; ratios of NSDUH to NESARC prevalence rates ranged from 2.1 to 5.7 percent for the illegal drug use outcomes. In NSDUH, past year substance use disorder (SUD) prevalence estimates were higher for cocaine and heroin, but they were similar to NESARC estimates for alcohol, marijuana, and hallucinogens. However, prevalence estimates for past year SUD conditional on past year drug use were substantially lower in NSDUH for marijuana, hallucinogens, and cocaine. Associations among drug use and SUD outcomes were substantially higher in NESARC. Total SUD prevalence did not differ between the two surveys, but estimates for blacks and Hispanics were higher in NSDUH. The authors believed that a number of methodological variables might have contributed to such discrepancies; among plausible candidates were factors related to privacy and anonymity, which may have resulted in higher drug use estimates in NSDUH, and differences in SUD diagnostic instrumentation, which may have resulted in higher SUD prevalence among past year substance users in NESARC.

### **Comparing drug testing and self-report of drug use among youths and young adults in the general population**

**CITATION:** Harrison, L. D., Martin, S. S., Enev, T., & Harrington, D. (2007). *Comparing drug testing and self-report of drug use among youths and young adults in the general population* (HHS Publication No. SMA 07-4249, Methodology Series M-7). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

**PURPOSE/OVERVIEW:** This report presents the results of a Validity Study conducted in 2000 and 2001 in conjunction with the National Household Survey on Drug Abuse (NHSDA), an annual survey to track the prevalence of substance use in the United States. The purpose of the Validity Study was to provide information on the validity of self-reported drug use in a general population survey by comparing the self-reports of respondents with the results of drug tests of urine and hair specimens obtained from those same respondents.

**METHODS:** The authors used the data from the Validity Study that was conducted as a supplement to the 2000 and 2001 NHSDAs. A separate national (excluding Alaska and Hawaii) sample of almost 6,000 individuals aged 12 to 25 was selected for the Validity Study, and more than 4,400 individuals completed an interview. These respondents were interviewed using the NHSDA methodology, with a slightly altered questionnaire to eliminate questions not needed for the study and to obtain key information needed for the study. Some of the questionnaire changes included adding questions about drug use corresponding to shorter time periods to be comparable with the window of detection of most drugs in urine and hair. In addition, a persuasion experiment was embedded in the study where half of the respondents were given a statement emphasizing the importance of accurate reporting before other questions specific to the Validity Study were asked. At the end of each interview, respondents were asked to provide a hair and a urine specimen, with an incentive of \$25 for each specimen. Exactly 4,000 respondents provided at least one specimen. Specimens were mailed to a testing laboratory, which conducted the drug tests and sent the results to the study team at the University of Delaware. Urine and hair specimens were screened using an immunoassay test for the following drug classes: marijuana/hashish (cannabinoids), cocaine, amphetamines, and opiates. Urine specimens also were analyzed for the presence of cotinine, the principal metabolite of nicotine, using only an

immunoassay test. For urine specimens, self-reported drug use (i.e., 30-day, 7-day, 3-day) was compared with confirmatory test results (i.e., positive or negative using the Validity Study cutoffs). Self-reported tobacco use (i.e., 30-day, 7-day, 3-day) was compared with the results of the urine test for cotinine (i.e., positive or negative using a 100 nanograms per milliliter [ng/mL] cutoff). Logistic regression models were developed to determine the correlates of overreporting and underreporting because logistic regression allows other variables that may affect the relationship between self-reports and drug test results to be controlled. After extensive bivariate and multivariate analyses were conducted, a set of variables was derived and used consistently to examine both underreporting and overreporting. These variables were gender, race (white, black, other), region of the country, religiosity, the privacy of the interview, whether the respondent received the experimental appeal to be truthful, difficulty remembering or understanding drug questions, truthfulness, friends' smoking, and passive exposure to the drug.

**RESULTS/CONCLUSIONS:** The Validity Study demonstrated its possibility to collect urine and hair specimens with a high response rate from individuals aged 12 to 25 in a household survey environment. The results of tests conducted on hair collected in this study could not be used to compare with self-reports because there were technical and statistical problems related to the hair tests and unresolved issues concerning the interpretation of the analytical results. Most youths aged 12 to 17 and young adults aged 18 to 25 reported their recent drug use accurately. However, there were some reporting differences in either direction—with some not reporting use and testing positive, and some reporting use and testing negative. Biological drug test results can be used as objective markers of drug use to verify self-reports.

### **Comparing the coverage of a household sampling frame based on mailing addresses to a frame based on field enumeration**

**CITATION:** Iannacchione, V., Morton, K., McMichael, J., Cunningham, D., Cajka, J., & Chromy, J. (2007). Comparing the coverage of a household sampling frame based on mailing addresses to a frame based on field enumeration. In *Proceedings of the 2007 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Salt Lake City, UT* (pp. 3323-3332). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** Cost savings, timeliness, and geographic diversity are primary advantages of using mailing addresses instead of field enumeration as a sampling frame for household surveys. The question is whether the advantages of mailing addresses are accompanied by a decrease in the coverage of the household population.

**METHODS:** The research was based on a probability sample of 50 segments that were assembled from census blocks in North Carolina. Within the geographic confines of each segment, the authors constructed two frames: one based on locatable residential mailing addresses and the other based on field enumeration. The authors used Global Positioning System technology to match the housing units (HUs) from each frame, without presuming that either approach is the "gold standard."

**RESULTS/CONCLUSIONS:** Overall, the authors found that field enumeration included approximately 98 percent of the HUs compared with 82 percent coverage for mailing addresses.

When restricted to occupied HUs, however, the coverage increased to approximately 99 and 95 percent, respectively. Equal coverage was found in 59 percent of occupied HUs in urban areas. In rural areas, however, mailing addresses were found to have significantly lower coverage than field enumeration. Locatable mailing addresses were nonexistent for 0.4 percent of HUs in areas without home delivery of mail. The authors estimated that field enumeration combined with the half-open interval frame supplementation methodology would yield virtually complete coverage of occupied and unoccupied HUs. An analogous methodology based on a letter carrier's delivery sequence would increase the coverage of locatable mailing addresses by at least 3.4 percentage points.

### **Evaluation of the effects of new noncore drug data on prevalence estimates in the National Survey on Drug Use and Health (NSDUH)**

**CITATION:** Kroutil, L., Vorburger, M., & Aldworth, J. (2007, October). Evaluation of the effects of new noncore drug data on prevalence estimates in the National Survey on Drug Use and Health (NSDUH). In *2006 National Survey on Drug Use and Health: Methodological resource book* (Section 18, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 39, RTI 0209009). Research Triangle Park, NC: RTI International.

**PURPOSE/OVERVIEW:** New questions were added to the 2006 National Survey on Drug Use and Health (NSDUH) to capture information about the use of drugs that respondents were not directly asked about in "core" sections of the interview. These new questions appeared in the noncore special drugs module, and this report presents findings from an investigation of the effects of these questions on drug use prevalence estimates in NSDUH. The specific new questions added in the noncore special drug modules were GHB (gamma hydroxybutyrate), also called "G," "Georgia Home Boy," "Grievous Bodily Harm," or "Liquid G"; Adderall<sup>®</sup> (a prescription stimulant); Ambien<sup>®</sup> (a prescription sedative); over-the-counter (OTC) cough or cold medicines; Ketamine (a hallucinogen), also called "Special K" or "Super K"; the tryptamine hallucinogens DMT (dimethyltryptamine), AMT (alpha-methyltryptamine), or Foxy (5-MeO-DIPT [5-methoxy-diisopropyltryptamine]); or *Salvia divinorum* (a hallucinogen).

**METHODS:** Measures of lifetime, past year, and past month use (or nonmedical use) of hallucinogens, stimulants, sedatives, prescription psychotherapeutic drugs, illicit drugs, and illicit drugs excluding marijuana that were based only on core data from the 2006 NSDUH were taken from the 2006 analytic file. The core variables that were used to construct these measures had been fully imputed to eliminate missing data. These measures were created in a manner consistent with the procedures in prior survey years. New measures also were created for the lifetime, past year, and past month based on the core drug use data and the noncore data from the new drug items in the special drugs module. Unlike the procedures for the core drug use measures described above, the noncore special drugs variables had undergone logical editing, but had not been statistically imputed. Consequently, these variables had missing data. For this study, missing data for the new drug variables were treated as being equivalent to lifetime nonuse or nonuse in the period of interest.

**RESULTS/CONCLUSIONS:** The inclusion of both core and noncore data from the 2006 NSDUH special drugs module for GHB, Adderall<sup>®</sup>, Ambien<sup>®</sup>, and hallucinogens had little effect on prevalence estimates for the use of hallucinogens, illicit drugs, and illicit drugs excluding marijuana compared with estimates based on core data alone. Inclusion of noncore data for Adderall<sup>®</sup> and Ambien<sup>®</sup> had some effect on estimates of the nonmedical use of prescription psychotherapeutic drugs, particularly for the lifetime period. In addition, inclusion of Adderall<sup>®</sup> and Ambien<sup>®</sup> data increased the estimate of past year nonmedical use of psychotherapeutic drugs by more than one percentage point for young adults aged 18 to 25. In contrast, inclusion of Adderall<sup>®</sup> and Ambien<sup>®</sup> data had notable effects on prevalence estimates for the more proximal measures of nonmedical use of stimulants and sedatives, respectively. In particular, many of the estimates of past year and past month misuse of sedatives more than doubled when estimates included both Ambien<sup>®</sup> data and core sedative data. Fewer than 1 million adults were estimated to be past year nonmedical users of sedatives based on core data alone compared with nearly 2.4 million based on core and Ambien<sup>®</sup> data.

### **Improving the sensitivity of needs assessment for substance abuse prevention planning: The measurement of differential severity of consequences for individual substance types**

**CITATION:** Shamblen, S. R., & Springer, J. F. (2007). Improving the sensitivity of needs assessment for substance abuse prevention planning: The measurement of differential severity of consequences for individual substance types. *Journal of Drug Education*, 37(3), 295-316.

**PURPOSE/OVERVIEW:** There is an absence of systematic, comparative research examining the negative consequences that are experienced as a result of using specific substances. Further, techniques typically used for needs assessment (i.e., prevalence proportions) do not take into account the probability of experiencing a negative consequence as a result of using specific substances.

**METHODS:** An approximated severity index was proposed that (a) takes into account the probability of experiencing negative consequences as a result of using specific substances and (b) is comparable across substances. Data from the National Survey on Drug Use and Health (NSDUH) and the Alcohol and Drug Services Study (ADSS) were used to demonstrate these techniques. The authors defined severity as the intersection of three conditions. First, an individual experiencing a severe consequence must have used the specific substance during his or her lifetime, indicating consumption of the substance. Second, the individual must have been in treatment during his or her lifetime, indicating a recognized dependence or problem-use condition. Third, the individual must have experienced specified negative consequences. When the number in a population within this subset was divided with the number of lifetime users in the denominator, the resulting probability was a severity score directly comparable across substances. It was the probability of experiencing a negative consequence in one's lifetime and having been in treatment in one's lifetime, given that one used the specified substance in one's lifetime.

**RESULTS/CONCLUSIONS:** The findings suggested that substances typically considered priorities based on prevalence proportions were not the same substances that had a high probability of causing negative consequences.

### **Understanding the relative influence of neighborhood, family, and youth on adolescent drug use**

**CITATION:** Wright, D. A., Bobashev, G., & Folsom, R. (2007). Understanding the relative influence of neighborhood, family, and youth on adolescent drug use. *Substance Use & Misuse*, 42(14), 2159-2171.

**PURPOSE/OVERVIEW:** A variety of programs have been developed to prevent substance use among youths in the United States. Oftentimes, these programs target youths directly and may have components that address the relational influence of families, schools, and communities. This article presents some new uses of variance components in understanding the influence of the neighborhood and family on individual drug use. In fields such as education, the use of variance components based on a nested hierarchical structure is more common, although it is rarely used in the area of substance use.

**METHODS:** The paper focused on marijuana use among youths, and the data came from the 1999 National Household Survey on Drug Abuse (NHSDA), which included a sample of approximately 25,000 youths aged 12 to 17. Using the data, the authors calculated variance components for a binary variable (0 or 1 for "yes" or "no"), which were similar to the methods discussed by Goldstein and Rasbash (1996). Specifically, they used several outcome variables for analysis, one being a direct measure of marijuana use and two others that were indirectly related to drug use.

**RESULTS/CONCLUSIONS:** The authors estimated variance components when the outcome was dichotomous and found that, for the use of marijuana in the past year, the role of the individual-level variables (individual adolescent vs. role of household vs. role of neighborhood) was quite prominent (79 percent of variation). A similar result was observed for the continuous scale variable of individual positive attitudes toward drug use (83 percent). For continuous constructs related to either the household level (parental monitoring) or the neighborhood level (neighborhood disorganization), the majority of variation still occurred at the individual level (67 and 51 percent, respectively), although they revealed significant percent variation (about 30 percent) at the corresponding family or neighborhood levels as well.

*This page intentionally left blank*

## 2008

### **Prevalence of nonmedical methamphetamine use in the United States**

CITATION: Durell, T. M., Kroutil, L. A., Crits-Christoph, P., Barchha, N., & Van Brunt, D. L. (2008). Prevalence of nonmedical methamphetamine use in the United States. *Substance Abuse Treatment, Prevention, and Policy*, 3, 19. doi:10.1186/1747-597X-3-

**PURPOSE/OVERVIEW:** Illicit methamphetamine use continues to be a public health concern in the United States. The goal of the current study was to use a relatively inexpensive methodology to examine the prevalence and demographic correlates of nonmedical methamphetamine use in the United States.

**METHODS:** The sample was obtained through an Internet survey of noninstitutionalized adults ( $N = 4,297$ ) aged 18 to 49 in the United States in 2005. Propensity weighting methods using information from the U.S. Census and the 2003 National Survey on Drug Use and Health (NSDUH) were used to estimate national-level prevalence rates. A two-stage weighting process was used to correct for possible selection bias within the Internet panel. First, the data were weighted according to results from a probability-based telephone survey using a propensity scoring approach. The propensity score adjusted for self-selection into the online population and into the panel and for survey nonresponse that may not be explained by demographic differences. The propensity score model was created by Harris using data from parallel telephone and Internet surveys that they periodically collect with the telephone survey based on random-digit dialing (RDD) probability sampling. Weights were created so that the weighted distribution of the propensity score for Internet respondents was matched to the distribution for the RDD telephone respondents. The second step in the weighting process was to weight the data to match the U.S. target population distribution by general demographic characteristics, as well as distribution of past month cigarette use and past month binge alcohol use (i.e., consumption of five or more drinks in a single occasion at least once in the past 30 days) estimated from the 2003 NSDUH, the most current publicly available data at the time the study was conducted.

**RESULTS/CONCLUSIONS:** The overall prevalence of current nonmedical methamphetamine use was estimated to be 0.27 percent. Lifetime use was estimated to be 8.6 percent. Current use rates for men (0.32 percent) and women (0.23 percent) did not differ, although men had a higher 3-year prevalence rate (3.1 percent) than women (1.1 percent). Within the age subgroup with the highest overall methamphetamine use (18 to 25 year olds), nonstudents had substantially higher methamphetamine use (0.85 percent current; 2.4 percent past year) than students (0.23 percent current; 0.79 percent past year). Methamphetamine use was not constrained to those with publicly funded health care insurance. Through the use of an Internet panel weighted to reflect U.S. population norms, the estimated lifetime prevalence of methamphetamine use among 18 to 49 year olds was 8.6 percent. These findings provided rates of use comparable with those reported in the 2005 NSDUH. Internet surveys are a relatively inexpensive way to provide complementary data to telephone or in-person interviews.

## **A clinical validation of the National Survey on Drug Use and Health assessment of substance use disorders**

CITATION: Jordan, B. K., Karg, R. S., Batts, K. R., Epstein, J. F., & Wiesen, C. (2008). A clinical validation of the National Survey on Drug Use and Health assessment of substance use disorders. *Addictive Behaviors*, 33(6), 782-798.

PURPOSE/OVERVIEW: Alcohol and illicit drug abuse and dependence continue to be of great national concern in the United States, as is true in other nations. The National Survey on Drug Use and Health (NSDUH) provides national annual estimates of substance use and abuse/dependence among the U.S. civilian, noninstitutionalized population aged 12 years or older. The authors conducted a clinical validation study of the substance use disorder questions of the NSDUH instrument using a sample of 288 adults and adolescents recruited from the community and outpatient substance abuse treatment programs in North Carolina.

METHODS: Using the Structured Clinical Interview for DSM-IV (SCID-IV) for adults and the Pittsburgh Adolescent Alcohol Research Center's Structured Clinical Interview (PAARC-SCID) for adolescents, the authors computed the psychometric properties of the NSDUH questions.

RESULTS/CONCLUSIONS: The authors found the level of agreement between the NSDUH and the SCID/PAARC-SCID interviews to be fair to moderate overall. There was somewhat better agreement for dependence than for abuse and for adults than for adolescents.

## **2006 and 2007 National Surveys on Drug Use and Health: Hallucinogen, methamphetamine, and prescription drug analysis**

CITATION: Kroutil, L., Davis, T., Handley, W., & Copello, E. (2008, September). 2006 and 2007 National Surveys on Drug Use and Health: Hallucinogen, methamphetamine, and prescription drug analysis. In *2007 National Survey on Drug Use and Health: Methodological resource book* (Section 16, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 39, RTI 0209009). Research Triangle Park, NC: RTI International.

PURPOSE/OVERVIEW: This study built on prior methodological research for the 2005 and 2006 National Surveys on Drug Use and Health (NSDUHs) that focused on the effects on prevalence estimates related to additional noncore questions about the use or nonmedical use of the following drugs: methamphetamine; Adderall<sup>®</sup> (a prescription stimulant); Ambien<sup>®</sup> (a prescription sedative); the hallucinogens ketamine, dimethyltryptamine (DMT), alpha-methyltryptamine (AMT), N, N-diisopropyl-5-methoxytryptamine (5 MeO-DIPT, also known as "Foxy"), and *Salvia divinorum*; and the illicit drug gamma hydroxybutyrate (GHB).

METHODS: Three types of drug use measures were developed for this report: lifetime, past year, and past month prevalence of drug use; initiation of methamphetamine use; and frequency of methamphetamine use in the past 12 months. These measures were based on core drug data alone or on core data plus relevant noncore data from the special drugs module.

**RESULTS/CONCLUSIONS:** In general, inclusion of both core and noncore data from the NSDUH special drugs module in 2006 and 2007 for methamphetamine, GHB, Adderall<sup>®</sup>, Ambien<sup>®</sup>, and hallucinogens did not affect most conclusions about trends in prevalence estimates between 2006 and 2007 for the measures affected by these drugs compared with published estimates based on core data alone (or core data plus noncore methamphetamine data, in the case of stimulants and psychotherapeutic drugs). However, data on nonmedical use of the prescription stimulant Adderall<sup>®</sup> appeared to affect the conclusions that would be reached about trends in nonmedical use of prescription stimulants or methamphetamine among young adults aged 18 to 25 and among males aged 12 or older. In particular, the past year and current prevalences of nonmedical stimulant use based on core stimulant data and noncore methamphetamine showed significant declines between 2006 and 2007 for young adults and males overall. Inclusion of data for nonmedical use of Adderall<sup>®</sup> yielded estimates of past year and current use for young adults and males that were no longer significantly different between these 2 years. In addition, inclusion of noncore data affected trends in the lifetime and past year prevalences of use of illicit drugs other than marijuana for youths aged 12 to 17. The core data showed significant declines between 2006 and 2007, but the core-plus-noncore data did not.

### **A robust procedure to supplement the coverage of address-based sampling frames for household surveys**

**CITATION:** McMichael, J. P., Ridenhour, J. L., & Shook-Sa, B. E. (2008). A robust procedure to supplement the coverage of address-based sampling frames for household surveys. In *Proceedings of the 2008 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Denver, Colorado* (pp. 4329-4335). Washington, DC: American Statistical Association.

**PURPOSE/OVERVIEW:** At the time of this study, address-based sampling (ABS) was becoming commonly used as an alternative to traditional methods, such as field enumeration, for household surveys because of its cost savings, timeliness, and geographic diversity. However, the coverage of ABS frames was not complete, and there was a need for both frame supplementation methods and evaluation of these methods. RTI developed a new procedure called the Check for Housing Units Missed (CHUM), and the authors discussed the implementation and testing of the procedures.

**METHODS:** The CHUM method uses three components to achieve complete coverage. Component 1 systematically identifies an address that is missing from the ABS frame. Component 2 identifies "missed areas" where component 1 will have no effect because an area's dwelling units (DUs) are not on the ABS frame. Component 3 identifies new streets that could interfere with the coverage properties of component 2.

**RESULTS/CONCLUSIONS:** The CHUM procedure was shown to be a robust method for supplementing coverage on an ABS study. It theoretically provided 100 percent coverage. However, the evaluation of this method used data from a study with relatively small geographic segments, and there may be new challenges with implementing this method in a study with larger geographic areas.

## **Barriers to survey participation among older adults in the National Survey on Drug Use and Health: The importance of establishing trust**

CITATION: Murphy, J., Schwerin, M., Eyerman, J., & Kennet, J. (2008). Barriers to survey participation among older adults in the National Survey on Drug Use and Health: The importance of establishing trust. *Survey Practice*, 1(2), 1-6.

PURPOSE/OVERVIEW: The U.S. Census Bureau estimated that 37 percent of the population will be aged 50 or older by 2030, up from about 28 percent in 2000. Accurate behavioral measures of these older Americans are vital to the making of sound, evidence-based policy decisions. A negative relationship between response rate and age has been identified in both survey methods and gerontological research, and the National Survey on Drug Use and Health (NSDUH) has shown that older individuals sampled in the survey, particularly those aged 50 or older, were less likely to complete the survey than younger individuals. This article discussed potential barriers to participation among this population group and practices that could improve response rates with older respondents.

METHODS: Twelve focus groups were conducted in three cities to explore the issue of nonresponse among potential respondents aged 50 or older. Focus group participants were asked to watch a series of videotaped vignettes filmed from the perspective of a survey respondent. They were asked to discuss reasons for participating or not participating in the survey. Specific questions focused on (1) the importance of the study topic, (2) privacy and confidentiality, (3) their opinion of the \$30 incentive, (4) thoughts on aspects of the interviewer's approach and the handheld device that the interviewer was using to record information, and (5) study materials, such as the lead letter and the question and answer (Q&A) brochure.

RESULTS/CONCLUSIONS: Focus group participants said they would be more likely to respond to a field interviewer (FI) who was prepared and polished, without being "slick." They expected FIs to perform their task in a professional manner—polite, positive, and knowledgeable of the survey questions. Concerns were raised about the survey approach and physical safety, fear of "scams," or other misuses of personal information. The importance of trusting the FI, the research organization, and the study purpose were expressed throughout all of the focus groups. Focus group participants reported that additional detailed information about the purpose and benefits of the study would facilitate trust and lend legitimacy to the research organization and the FI. Participants expressed confusion over the description of the selection process and the meaning of "random selection." In general, the offer of a \$30 incentive was not seen as persuasive by the focus group participants.

## **Using the K6 scale to screen for serious mental illness among criminal justice populations: Do psychiatric treatment indicators improve detection rates?**

CITATION: Swartz, J. A. (2008). Using the K6 scale to screen for serious mental illness among criminal justice populations: Do psychiatric treatment indicators improve detection rates? *International Journal of Mental Health and Addiction*, 6(1), 93-104. doi:10.1007/s11469-007-9107-3

**PURPOSE/OVERVIEW:** This study evaluated the accuracy of the Kessler-6 (K6), a screening tool for severe psychological distress associated with serious mental illness, with and without additional questions on past year psychiatric treatment.

**METHODS:** The authors used the Composite International Diagnostic Interview Short Form (CIDI-SF) as the gold standard and used for comparison the data from 3,534 men and 1,350 women who indicated past year criminal justice involvement from the 2001 National Household Survey on Drug Abuse (NHSDA) and the 2002 National Survey on Drug Use and Health (NSDUH).

**RESULTS/CONCLUSIONS:** The authors found the unmodified K6 to be about equally accurate for men and women with receiver operating characteristic (ROC) and area under the curve (AUC) values for both at .88 and sensitivity scores in the range of 62.4 to 74.0 for men and 70.3 to 80.9 for women. Inclusion of psychiatric treatment indicators increased sensitivity but decreased specificity of the K6 for both men and women and consequently did not improve overall performance. Based on these findings, the authors tentatively recommended use of the K6 with criminal justice populations. They also recommended more rigorous studies to establish an optimum threshold and to evaluate whether the K10 (the longer version of the K6) can improve on the sensitivity and positive predictive values obtained with the K6.

### **Intraclass correlation patterns of cognitive and behavioral measures of illicit drug use within six major metropolitan areas**

**CITATION:** Zhang, Z., Cohen, M. P., & Wright, D. (2008). Intraclass correlation patterns of cognitive and behavioral measures of illicit drug use within six major metropolitan areas. In *Proceedings of the 2008 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Denver, CO* (pp. 3986-3993). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** Estimating components of variance using large-scale data is useful because large numbers of clusters may be included, which can lead to more precise estimates of intraclass correlations (ICCs). Considering that the National Survey on Drug Use and Health (NSDUHs) is conducted at the national level, the subnational level of ICC estimations may enrich the understanding of the differentiated ICCs across metropolitan areas. The authors estimated ICCs for six metropolitan areas separately so that the results can better inform the localized community interventions.

**METHODS:** Using data from the National Household Survey on Drug Abuse (NHSDA, the prior name for NSDUH), the authors calculated ICCs for both cognitive and behavioral measures on drug use at the census tract and census block group levels within six major metropolitan areas, respectively.

**RESULTS/CONCLUSIONS:** ICCs and design effects, varying from variable to variable, are often large and cannot be ignored. ICC is more generalizable and preferred compared with the design effect or variance inflation factor because the latter are dependent on the clusters. The authors demonstrated that ICCs should not be overlooked in the substance use field and

discussed further the utility of empirical knowledge of ICCs of pertinent measures in future sample designs, estimations, and policy-oriented preventions and interventions.

## 2009

### **Screening for serious mental illness in the National Survey on Drug Use and Health (NSDUH)**

CITATION: Colpe, L. J., Epstein, J. F., Barker, P. R., & Gfroerer, J. C. (2009). Screening for serious mental illness in the National Survey on Drug Use and Health (NSDUH). *Annals of Epidemiology*, 19(3), 210-211. doi:10.1016/j.annepidem.2008.09.005

PURPOSE/OVERVIEW: Public Law (Pub. L.) No. 102-321, the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) Reorganization Act, created the Substance Abuse and Mental Health Services Administration (SAMHSA) and established a block grant program for U.S. States to fund community mental health services for adults with serious mental illness (SMI). The criteria for SMI require adults (18 years or older) to have at least one 12-month *Diagnostic and Statistical Manual of Mental Disorders* (DSM) disorder, other than a substance use disorder, resulting in functional impairment that substantially interferes with or limits one or more major life activities. Since then, national SMI estimates have been generated from mental disorder diagnosis and impairment data yielded from the National Comorbidity Survey (NCS) and, most recently, the National Comorbidity Survey Replication (NCS-R). Although the NCS-R is a large, in-depth psychiatric epidemiology study, it is not conducted frequently enough to produce annual or biennial estimates, nor is it designed to generate State-level estimates. To address these data needs, SAMHSA's National Survey on Drug Use and Health (NSDUH) was nominated as the vehicle to collect adult SMI prevalence estimates. A methodological study to examine the accuracy of a variety of psychiatric symptom and impairment screeners to predict SMI in NSDUH was conducted in 2000.

METHODS: The study obtained clinical assessments on a small sample of NSDUH respondents in the Boston area. All of the scales used in the methodology study were included in a new mental health module added to the 2001 NSDUH. In 2004, a separate module to estimate the lifetime and 12-month prevalence of major depressive episode (MDE) was added to the NSDUH questionnaire. A split-sample experiment was embedded within the 2004 NSDUH to measure how the removal of the extra scales would affect overall SMI estimates. Half of the adult sample ( $n = 22,628$ ) was administered the mental health module used in the 2003 NSDUH questionnaire, without the MDE questions but with all of the screeners that were to be dropped. The other half of the adult sample ( $n = 22,825$ ) was administered the module that had been revised for the 2004 NSDUH, with the MDE section placed after the Kessler-6 (K6, a mental disorder scale) and the extra scales removed.

RESULTS/CONCLUSIONS: The results of the experiment showed large differences between the two samples in both the K6 total score and the proportion of respondents with a K6 score above the SMI cut point of 12. These differences suggested that the K6 scale was context-dependent; that is, respondents appeared to respond to the K6 items differently depending on whether or not the scale was preceded by a broad array of other mental health questions. These findings led SAMHSA to launch an additional study to calibrate K6 and impairment data obtained from NSDUH respondents against a gold-standard clinical assessment in a nationally

representative sample. In the interim, SAMHSA collected and reported estimates of serious psychological distress (based on K6 scores) and MDE among the U.S. adult population and continued to do so while working to develop an improved measure of SMI.

## **Redesigning the National Survey on Drug Use and Health**

CITATION: Gfroerer, J., Bose, J., Painter, D., Jones, M., & Kennet, J. (2009, November). *Redesigning the National Survey on Drug Use and Health*. Paper presented at the 2009 Federal Committee on Statistical Methodology Research Conference, Washington, DC.

PURPOSE/OVERVIEW: The National Survey on Drug Use and Health (NSDUH) is an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years old or older. The survey is used to produce national and State-level estimates of the prevalence of use of illicit drugs, alcohol, and tobacco products as well as measures related to mental health. This paper describes the redesign process and the challenges faced in achieving the following three goals: (1) maintain valid trend measurement, (2) update and improve the questionnaire and survey methodology, and (3) keep costs within expected budget levels.

METHODS: N/A.

RESULTS/CONCLUSIONS: The authors listed and summarized seven diverse kinds of studies related to the NSDUH redesign: (1) prior research studies relevant for the NSDUH redesign, (2) new research related to sampling, (3) new studies related to improving response rates, (4) new studies assessing NSDUH estimation methods, (5) developing a new questionnaire, (6) other new studies related to the NSDUH redesign, and (7) implementation issues. The authors then shared a number of lessons learned from 2005 to date while planning and preparing for the 2013 NSDUH redesign, including the following: look to the past, look sideways, look ahead, consult with data users, keep management in the loop, approach the redesign in its entirety, look for interconnections, keep key and secondary outcome measures in mind, trust field tests—to a point, trust expert knowledge, think like a fed, and anticipate failures.

# 2010

## **The National Survey on Drug Use and Health Mental Health Surveillance Study: Calibration analysis**

CITATION: Aldworth, J., Colpe, L. J., Gfroerer, J. C., Novak, S. P., Chromy, J. R., Barker, P. R., Barnett-Walker, K., Karg, R. S., Morton, K. B., & Spagnola, K. (2010). The National Survey on Drug Use and Health Mental Health Surveillance Study: Calibration analysis. *International Journal of Methods in Psychiatric Research*, 19(Suppl. 1), 61-87. doi:10.1002/mpr.312

**PURPOSE/OVERVIEW:** The Mental Health Surveillance Study (MHSS) was an initiative by the Substance Abuse and Mental Health Services Administration to develop and implement methods for measuring the prevalence of serious mental illness (SMI) among U.S. adults aged 18 or older. The 2008 MHSS used data from clinical interviews that were administered to a subsample of respondents from the National Survey on Drug Use and Health (NSDUH) to accurately calibrate mental health screening scale data for estimating the prevalence of SMI in the full NSDUH sample.

**METHODS:** This research was based on the mental health scales that were included in the Kessler-6 (K6) screening scale of psychological distress (administered to all respondents) along with two additional measures of functional impairment (each administered to a random half sample of respondents): the World Health Organization Disability Assessment Schedule (WHODAS) and the Sheehan Disability Scale (SDS). The Structured Clinical Interview for DSM-IV (SCID) was administered to a subsample of 1,506 adult NSDUH respondents within 4 weeks of completing the NSDUH interview. Descriptive analyses were conducted to examine the distribution of respondent characteristics in the MHSS to check for imbalances between the two half samples, each of which was assigned to one of the impairment scales. Modeling analyses were conducted to develop algorithms based on the K6 scale and each of the impairment scales in turn, with the goal of identifying the best possible model for each impairment scale. This involved fitting a variety of models using alternative predictors, including different forms of the K6 and impairment variables. For each model, receiver operating characteristic (ROC) analyses were conducted to select the optimal cut point for determining SMI status. Weighted counts were used in the ROC classifications because primary interest is in estimating SMI status in the adult U.S. population. Models to determine SMI were compared and evaluated based on three criteria: (1) model robustness (e.g., preference given to parsimonious models that could be generalized to data beyond that used in the modeling process); (2) minimization of misclassification errors in SMI prediction (i.e., exhibiting reasonable ROC statistics, such as sensitivity and AUC, defined as the area under the ROC curve based on the optimal cut point described above); and (3) reasonable SMI estimates based on the full dataset (i.e., balanced across several demographic subgroups and across the WHODAS and the SDS half samples).

**RESULTS/CONCLUSIONS:** The authors found that the models with the WHODAS were more robust, while SMI prediction accuracy of the K6 was improved by adding either the WHODAS

or the SDS to the prediction equation. The results of the calibration study and methods used to derive prevalence estimates of SMI were also presented.

### **Use of single years of age in the National Survey on Drug Use and Health (NSDUH) weighting to improve drug prevalence estimates**

CITATION: Chen, P., Sathe, N., Jones, M., Dai, L., Laufenberg, J., Folsom, R. E., & Gordek, H. (2010). Use of single years of age in the National Survey on Drug Use and Health (NSDUH) weighting to improve drug prevalence estimates. In *Proceedings of the 2010 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Vancouver, British Columbia, Canada* (pp. 3055-3066). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: In the National Survey on Drug Use and Health (NSDUH), several age groups (i.e., 12 to 17, 18 to 25, 26 to 34, 35 to 49, 50 to 64, and 65 or older) are currently used in the nonresponse (NR) and poststratification (PS) adjustments. Using 2004 to 2006 NSDUH data, the authors found that the response rate decreased as the age increased in the 12 to 25 age range. Additionally, prevalence rates for illicit drug, alcohol, and tobacco use increased almost linearly between the ages of 12 and 21, reaching a peak at age 21. Because the response rates and drug use prevalence rates changed dramatically between the ages of 12 and 21, use of single years of age instead of age groups for both the NR and PS adjustments should reduce NR bias and the variance of estimates. This paper explores the use of single years of age between 12 and 25 in the 2006 NSDUH weighting process and discusses the prevalence rates and standard errors produced using the new set of weights.

METHODS: The authors adopted the following approach for using single years of age as predictors in the person-level NR and PS adjustments for each of the nine census division-level model groups: (1) added single years of age in the main effect (12, ..., 25) for both person-level NR and PS in place of the 12 to 17 and the 18 to 25 age groups; (2) kept all of the other variables in the models; (3) used age groups 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older in the interactions for NR; and (4) used age groups 12 to 17, 18 to 25, 26 to 34, 35 to 49, 50 to 64, and 65 or older in the interactions for PS. After recalibrating the weights using single years of age as the main effects in the NR and PS adjustments, the authors checked the distribution of the recalibrated weights against the current analysis weights that used age groups as the main effects in the generalized exponential models. Then they examined the estimated numbers of past year users and the prevalence rates for a selected set of outcomes using the recalibrated weights (which used single year of age variables in the NR and PS adjustments) and compared them with the estimates based on the current weights (without the single year of age variables).

RESULTS/CONCLUSIONS: Adding single years of age (12 to 25) in the person-level NR and PS adjustments did not change the weight distribution dramatically. Differences between the estimated numbers and percentages of users for the younger age groups and single years of age (12 to 25) using the two sets of weights were observed, while the impact for other demographic domains was minimal. The effect on the estimated numbers of users (counts) was greater than on the estimated percentages for the 12 to 17 and the 18 to 25 age groups and for the single years of age. There was also evidence that the recalibration with single years of age (12 to 25) reduced

errors (lower mean square error using the recalibrated weights) for individuals aged 12 to 25 in the NSDUH sample.

## **Reliability of key measures in the National Survey on Drug Use and Health**

CITATION: Chromy, J. R., Feder, M., Gfroerer, J., Hirsch, E., Kennet, J., Morton, K. B., Piper, L., Riggsbee, B. H., Snodgrass, J. A., Virag, T. G., & Yu, F. (2010, February). *Reliability of key measures in the National Survey on Drug Use and Health* (Methodology Series M-8, HHS Publication No. SMA 09-4425). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

PURPOSE/OVERVIEW: Information on data quality is an important output of major Federal surveys because survey data often are used to influence policy decisions. A range of types of error may occur in surveys related to problems with the respondent's understanding of the questions or the effects of the interviewer. These problems may bring about a disparity between the survey response and a true value. Reinterviewing survey respondents in studies of survey response reliability provides a direct measure of such response variance. As a response to a 2006 directive by the Federal Government's Office of Management and Budget, a reinterview study of respondents to the National Survey on Drug Use and Health (NSDUH) was conducted.

METHODS: The Reliability Study was embedded within the 2006 NSDUH main study. A subsample of the main study sample of 67,802 was selected such that data from the initial interview were used for both the main study and the Reliability Study. As for the main study, the respondent universe included the civilian, noninstitutionalized population aged 12 or older. In the Reliability Study, the respondent universe excluded residents of Alaska and Hawaii, residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories), and individuals who did not speak English. To preserve the results and response rates of the main study, neither the field interviewers (FIs) nor respondents were informed ahead of time about the selection for the second interview. Recruitment scripts for the second interview were added to the end of the first interview and administered to the 3,516 eligible respondents selected for reinterview. Second interviews were obtained from 3,136 respondents, for an 85.6 percent reinterview weighted response rate. Although an incentive of \$30 was offered for the first interview, \$50 was offered for participation in the second interview. The second interviews were conducted 5 to 15 days following the initial interview with the same questionnaire as for the first interview. A set of follow-up questions was added to the end of the second interview about respondent use of tobacco, alcohol, and other drugs between the two interviews. A substudy was conducted within the Reliability Study—the same versus different interviewer substudy—to examine the potential impact that an FI might have on reliability.

RESULTS/CONCLUSIONS: Responses for substance use in the lifetime had almost perfect reliability, and responses for substance use in the past year showed substantial agreement. Age at first use of specific substances showed mostly moderate reliability, but findings for which a substance was used first were less consistent, with some being of only fair reliability. The reliability of responses to age at last use was generally fair. Comparisons of the consistency of responses among those who were interviewed by the same versus different FIs at the time of the two interviews showed no significant effect of the interviewer on the reliability of survey

responses. The consistency of responses among those whose first and second interviews were fewer than 9 days apart was similar to the consistency of responses among those whose interviews were 9 or more days apart. Analyses showed that questions about factual personal events or characteristics were more reliable than questions that asked for a respondent's personal opinion or intentions or questions that addressed issues that involved perceived discrimination (i.e., carried a social stigma).

## **The National Survey on Drug Use and Health Mental Health Surveillance Study: Calibration study design and field procedures**

CITATION: Colpe, L. J., Barker, P. R., Karg, R. S., Batts, K. R., Morton, K. B., Gfroerer, J. C., Stolzenberg, S. J., Cunningham, D. B., First, M. B., & Aldworth, J. (2010). The National Survey on Drug Use and Health Mental Health Surveillance Study: Calibration study design and field procedures. *International Journal of Methods in Psychiatric Research*, 19(Suppl. 1), 36-48. doi:10.1002/mpr.311

PURPOSE/OVERVIEW: The Mental Health Surveillance Study (MHSS) was an initiative by the Substance Abuse and Mental Health Services Administration (SAMHSA) to monitor the prevalence of serious mental illness (SMI) among adults in the United States. In 2008, the MHSS used data from clinical interviews to calibrate mental health data from the National Survey on Drug Use and Health (NSDUH) for estimating the prevalence of SMI based on the full NSDUH sample. This paper describes the MHSS calibration study procedures, including information on sample selection, instrumentation, follow-up, data quality protocols, and management of distressed respondents.

METHODS: The clinical interview used in this study was the Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV; SCID). NSDUH interviews were administered via audio computer-assisted self-interviewing (ACASI) to a nationally representative sample of the population aged 12 years or older. A total of 46,180 NSDUH interviews were completed with adults aged 18 years or older in 2008. The SCID was administered by mental health clinicians to a subsample of 1,506 adults via telephone. Clinical interviews were conducted by master's and doctoral level mental health professionals who had been carefully and extensively trained to administer the semistructured clinical interview over the telephone. The study protocol included comprehensive instructions for identifying and managing distressed respondents and for ongoing supervision and interrater training exercises for the clinical interviewers.

RESULTS/CONCLUSIONS: Descriptive analyses of the demographic characteristics of the clinical interview sample indicated that the sample was balanced and consistent with the overall NSDUH sample. A 76 percent unweighted completion rate among those who agreed to the clinical interview was achieved by the study, a commendable rate given the shortness of the 4-week data collection period and the lack of in-person follow-up. Given the success in the execution of the 2008 study, SAMHSA decided to continue to include the K6 and WHODAS scales in the main NSDUH interview and to collect clinical interview data from a subset of NSDUH respondents to monitor the prevalence of SMI among adults in the United States. It was decided that this will allow additional analysis of SMI at the State level, as well as investigations

into the prevalence and impact of milder forms of mental illness (e.g., with mild to moderate functional impairment). These continuing calibration activities and the ongoing nature of the study are significant contributions to mental health surveillance in the United States.

### **The influence of prior experiences in managing current and future risks during survey transition points on the National Survey on Drug Use and Health (NSDUH)**

CITATION: Gfroerer, J., & Bose, J. (2010). The influence of prior experiences in managing current and future risks during survey transition points on the National Survey on Drug Use and Health (NSDUH). In *Proceedings of the 2010 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Vancouver, British Columbia* (pp. 422-430). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: Since its inception in 1971, the National Survey on Drug Use and Health (NSDUH) has experienced many changes, including the transfer of the survey between Federal agencies; changes in government project officers and contractors; modifications to questionnaire content; major changes to the sample design and size; introduction of new modes of data collection and incentives to respondents; changes in the oversight and management of field staff; and introduction of new weighting, editing, and imputation methods. Some of these changes have, not surprisingly, resulted in both intended and unintended consequences, in some cases despite best efforts to control and quantify the effects of these changes. This paper uses examples to illustrate how prior experiences have influenced both ongoing practices and the current approach to redesigning the survey.

METHODS: N/A.

RESULTS/CONCLUSIONS: First, accurate trend measurement with an ongoing cross-sectional survey requires careful monitoring of data collection and estimation procedures to ensure comparability. Caution is needed when sampling errors are small. Small field tests cannot always be relied on to make decisions about design changes. Second, any major redesign of a large ongoing survey is probably going to result in a break in trends. Although it is probably possible to implement some improvements that have a low probability of disrupting the trend, and it may be feasible to implement a redesign under a split-sample design to account for and measure methods effects, there is no guarantee that this will be successful. Promises of trend continuation after a redesign are probably ill advised. The choice often comes down to maintaining the trend with a problematic design and biased estimates versus improving the survey (including possibly saving on costs) but breaking the trend.

### **The validity of self-reported tobacco and marijuana use, by race/ethnicity, gender, and age**

CITATION: Hughes, A., Heller, D., & Marsden, M. E. (2010, May). The validity of self-reported tobacco and marijuana use, by race/ethnicity, gender, and age. In L. A. Aday & M. Cynamon (Eds.), *Ninth Conference on Health Survey Research Methods: Conference*

*proceedings* (pp. 132-142). Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

**PURPOSE/OVERVIEW:** Researchers and policymakers have long been concerned about the validity of self-reported drug use and have provided recommendations for improvement, including the use of biological specimens to validate self-reports (U.S. General Accounting Office, 1993). A National Institutes of Health strategic plan for reducing health disparities in drug abuse and dependence recommended improving the validity of self-reported drug use because minority populations may be differentially affected (National Institute on Drug Abuse, 2004). The goal of this study was to examine the nature and extent of bias and discordance in self-reported estimates of tobacco and marijuana use by race/ethnicity, gender, and age compared with results from urinalysis tests in a nationally representative household survey.

**METHODS:** This paper used the data collected through the National Household Survey on Drug Abuse (NHSDA) Validity Study. The total number of respondents in the Validity Study was 4,465 over the 2-year data collection period (2000 and 2001), with a 74.3 percent weighted interview response rate. Of those completing the interview, 89.4 percent provided hair, urine, or both; of these, 80.5 percent provided both, 4.7 percent provided only urine, and 4.3 percent provided only hair. In addition to weighted prevalence rates of self-reported 3-day use and positive test results, statistics comparing self-reporting and drug testing also were calculated. These included estimates of discordance, underreporting and overreporting, bias, and correlation between discordance and bias. Two logistic regression models predicting past 3-day discordance of tobacco and marijuana were estimated. Models were fit using covariates including gender, age, race/ethnicity, passive exposure in the past 6 months to tobacco or marijuana, and socioeconomic status (SES).

**RESULTS/CONCLUSIONS:** All self-reported estimates of tobacco and marijuana use exhibited a downward bias, meaning that underreporting occurred more frequently than overreporting. Blacks and youths aged 12 to 14 had the largest bias compared with others in their respective demographic groups for both tobacco and marijuana use. Unadjusted odds ratios showed that older individuals and blacks were more likely than the youngest age group and whites, respectively, to report discrepant responses compared with urine test outcomes; moreover, females were less likely to misreport than males. However, after controlling on SES, privacy, truthfulness, friends' use, and other theoretically relevant covariates, these relationships were no longer statistically significant.

### **The best of both worlds: A sampling frame based on address-based sampling and field enumeration**

**CITATION:** Iannacchione, V., Morton, K., McMichael, J., Shook-Sa, B., Ridenhour, J., Stolzenberg, S., Bergeron, D., Chromy, J., & Hughes, A. (2010, August). *The best of both worlds: A sampling frame based on address-based sampling and field enumeration*. Presented at the 2010 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Vancouver, British Columbia.

**PURPOSE/OVERVIEW:** Cost savings are the primary advantage of using address-based sampling (ABS) over field enumeration (FE) for in-person surveys of the civilian, noninstitutionalized population. However, the values of cost savings decrease by research that indicates that FE provides more complete coverage than ABS, especially in rural areas. The authors developed and piloted a candidate sampling frame for the National Survey on Drug Use and Health (NSDUH) that uses ABS supplemented with a frame-linking procedure in area segments where they anticipate adequate ABS coverage and FE in segments where they anticipate poor ABS coverage. The objective of the candidate frame is to lower costs without sacrificing coverage levels of the current NSDUH sampling frame, which is based solely on FE.

**METHODS:** In the first Check for Housing Units Missed (CHUM1) procedure, field interviewers (FIs) start at an address that is contained on the ABS list (the starting dwelling unit [DU]) and follow a predetermined path of travel (usually moving to the left of the starting DU) until they either reach an address that is contained on the ABS list or they return to the starting DU. This allows coverage of DUs that are not on the ABS list but are in the same block as addresses on the ABS list. In the CHUM2 procedure, FIs perform the CHUM procedure from a predetermined start point in a randomly selected area rather than a starting DU. The FIs follow the same path of travel that they do for the CHUM1 procedure, stopping when they either list an address that matches to the ABS list or they return to the start point without finding a match. This procedure enables coverage of DUs that are in blocks where none of the DUs are on the ABS list.

**RESULTS/CONCLUSIONS:** The gains in coverage afforded by the CHUM procedure are essential to the veracity of the hybrid frame. Similar to the half-open interval (HOI) procedure, the CHUM procedure offers virtually complete coverage of DUs if implemented correctly in the field. The hybrid frame, utilizing a combination of FE, the ABS list, and the CHUM1 and CHUM2 procedures, theoretically provides 100 percent coverage of the target population. The authors report on the trade-offs between coverage and cost savings as area segments are shifted from FE to ABS.

## **Reliability and data quality in the National Survey on Drug Use and Health**

**CITATION:** Kennet, J., Gfroerer, J., Barker, P., Piper, L., Hirsch, E., Granger, R., & Chromy, J. R. (2010, May). Reliability and data quality in the National Survey on Drug Use and Health. In L. A. Aday & M. Cynamon (Eds.), *Ninth Conference on Health Survey Research Methods: Conference proceedings* (pp. 107-123) Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

**PURPOSE/OVERVIEW:** High reliability is a necessary condition to guarantee data validity. If a question or set of questions proves unreliable, discussion of disparities in the construct purportedly being measured loses its grounding. This paper presents preliminary results from a reliability study carried out on the National Survey on Drug Use and Health (NSDUH). More than 3,100 respondents participated in the study, and this rich dataset is likely to yield many important findings in the future.

**METHODS:** The data used in this paper came from the reliability study data obtained in the second and third quarters of 2006, which included about 2,200 respondents. The reliability study sample was drawn from the NSDUH main sample. For practical reasons, respondents in Alaska and Hawaii were not included in the reliability study, nor were non-English speaking respondents. Recruitment scripts were added to the end of the first interview to the eligible respondents selected for reinterview. A \$30 incentive was offered at the first interview, and \$50 was offered at the second interview. A set of follow-up questions was also added to the end of the second interview about respondent use of tobacco, alcohol, and other drugs between the two interviews.

**RESULTS/CONCLUSIONS:** Youths appeared less consistent than young adults and older adults in their reporting of substance use, particularly in the cases of lifetime and past year nonmedical prescription drug use and past year alcohol use. Young adults and older adults had fewer kappas below those of the complement of the sample and, in a few cases, appeared more consistent than the other groups. Other patterns that appeared were related to race/ethnicity. White respondents appeared to be more consistent than others in reporting past year substance use. A general pattern also appeared to be present in the case of income. Greater income was generally associated with greater response consistency. This result would only be surprising if educational level did not exhibit the same general pattern of association, which it did. These findings point toward the importance of efforts to lower the reading level of the instrument and to improve the comprehensibility of the supporting materials that, among other topics, describe the survey, the uses of the data, and measures taken to enhance confidentiality.

### **Estimated drug use based on direct questioning and open-ended questions: Responses in the 2006 National Survey on Drug Use and Health**

**CITATION:** Kroutil, L. A., Vorburger, M., Aldworth, J., & Colliver, J. D. (2010). Estimated drug use based on direct questioning and open-ended questions: Responses in the 2006 National Survey on Drug Use and Health. *International Journal of Methods in Psychiatric Research*, 19(2), 74-87. doi:10.1002/mpr.302

**PURPOSE/OVERVIEW:** Substance use surveys may use open-ended items to supplement questions about specific drugs and obtain more exhaustive information on illicit drug use. However, these questions are likely to underestimate the prevalence of use of specific drugs. Little is known about the extent of such underestimation or the groups most prone to underreporting.

**METHODS:** Using data from the 2006 National Survey on Drug Use and Health (NSDUH), a civilian, noninstitutionalized population survey of individuals aged 12 or older in the United States, the authors compared drug use estimates based on open-ended questions with estimates from a new set of direct questions that occurred later in the interview.

**RESULTS/CONCLUSIONS:** For these drugs, estimates of lifetime drug use based on open-ended questions often were at least 7 times lower than those based on direct questions. Among adults identified in direct questions as substance users, lower educational levels were consistently associated with nonreporting of use in the open-ended questions. Given NSDUH's

large annual sample size (approximately 67,000 interviews), combining data across future survey years could increase the understanding of characteristics associated with nonreporting of use in open-ended questions and allow drug use trends to be extrapolated to survey years in which only open-ended question data are available.

### **Questionnaire design considerations when expanding a survey target population to include children**

CITATION: LeBaron, P. A., Granger, R., Park, H., Heller, D., Dean, E., & Bettinger, G. (2010). Questionnaire design considerations when expanding a survey target population to include children. In *Proceedings of the 2010 Joint Statistical Meetings, American Statistical Association, AAPOR, Chicago, IL* (pp. 6259-6273). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: When designing a questionnaire, survey practitioners are challenged with developing questions that are appropriate for all respondents within the target population. This task becomes increasingly difficult when the same survey instrument will be administered to both children and adults. This paper examines factors to be considered when asking children to respond to the same survey items as adults.

METHODS: In order to evaluate the impact on questionnaire design and nonresponse of expanding a target population to include children under 12 years of age, the authors used data from the National Survey on Drug Use and Health (NSDUH). The authors examined differences in cognitive ability, nonresponse, and timing data between minors in the NSDUH sample. The differences in respondent behavior of the single years of age within the 12 to 17 year old age group provided insight into the effects of administering a survey to respondents under 12 years of age.

RESULTS/CONCLUSIONS: For expanding a survey target population to include children under 12 years of age, the authors suggested expanding the target population in such a way that would not require the development of new or significantly altered screening or interview instruments, such as only including children aged 10 or 11. This would potentially result in only minor adjustments to current data collection protocols while limiting the impact on data quality, instrumentation changes, and costs. Extensive usability testing and cognitive interviewing would need to be conducted to inform this decision and to ensure that children aged 10 or 11 could comprehend the questionnaire items and respond accurately to sensitive questions.

### **Predicting the coverage of address-based sampling frames prior to sample selection**

CITATION: McMichael, J. P., Ridenhour, J. L., Shook-Sa, B. E., Morton, K. B., & Iannacchione, V. G. (2010). Predicting the coverage of address-based sampling frames prior to sample selection. In *Proceedings of the 2010 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Vancouver, British Columbia* (pp. 4852-4859). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** The current sampling frame for the National Survey on Drug Use and Health (NSDUH) relies on field enumeration (FE) supplemented with the half-open interval (HOI) procedure (Kish, 1965). Because of the costs associated with FE, the size of the area segments is small—usually, about 100 dwelling units (DUs) in a rural area and 150 DUs in an urban area. Several national in-person surveys are looking at address-based sampling (ABS) instead of or in conjunction with FE due to the lower costs associated with ABS. In 2009, RTI conducted a field study for NSDUH aimed at investigating the cost implications and coverage properties of a sampling frame based on ABS in area segments with adequate ABS coverage and FE elsewhere. The objective of the NSDUH field study was to develop and test an ABS/FE hybrid frame that provides cost savings without sacrificing coverage.

**METHODS:** The field study was implemented by subsampling 200 NSDUH segments from the 2009 quarter 1 sample. The sample had 3,878 screened and eligible sampled DUs in a subsample of 200 NSDUH segments. Segments in Alaska and Hawaii were excluded from the field study sampling frame. To develop a hybrid frame of DUs, the authors attempted to match the addresses of eligible sampled DUs obtained from NSDUH's FE process and updated during the screening to a list of mailing addresses purchased from a commercial vendor. The authors classified the segments by coverage threshold to theoretically evaluate how these segments would be allocated to FE or ABS under the hybrid frame.

**RESULTS/CONCLUSIONS:** Under a hybrid frame, correctly predicting whether segments should utilize ABS or FE is essential to retaining desired coverage properties (e.g., minimizing coverage bias) while achieving cost savings. Cost savings are achieved by appropriately allocating segments that have sufficient ABS coverage to ABS. Coverage is maintained by allocating segments where ABS coverage is low to FE. Allocating correctly, in both cases, reduces costs.

### **Address-based sampling and the National Survey on Drug Use and Health: Evaluating the effects of coverage bias**

**CITATION:** Morton, K., McMichael, J. P., Ridenhour, J. L., & Bose, J. (2010). Address-based sampling and the National Survey on Drug Use and Health: Evaluating the effects of coverage bias. In *Proceedings of the 2010 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Vancouver, British Columbia* (pp. 4902-4907). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** A common concern of survey researchers is whether the coverage properties of an address-based sampling (ABS) frame create outcome bias in the estimates from in-person surveys. This paper evaluates basic demographics and several drug use and mental health measures obtained from 1,725 respondents in a probability sample of 200 area segments from the National Survey on Drug Use and Health (NSDUH).

**METHODS:** The evaluation compared outcomes from respondents covered by the NSDUH's field enumeration (FE) frame with those covered by an ABS frame derived from the United States Postal Service Computerized Delivery Sequence (USPS CDS) file. After poststratifying

the weights to populations with known ABS undercoverage, the authors tested for significant differences in outcomes between the two frames.

**RESULTS/CONCLUSIONS:** Estimates based on the ABS-only frame were limited to small, but statistically significant differences when compared with the FE frame. Use of the Check for Housing Units Missed (CHUM) frame-linking procedure to supplement the ABS frame helped to mitigate some of these differences. It is notable that these comparisons have the statistical power to declare very small differences in the overall prevalence estimates statistically significant because the estimates based on the FE frame and the estimates based on the ABS frame share a large portion of their cases. For example, for a prevalence estimate of 0.01 percent, a difference of 0.002 percent can be detected with 80 percent power and a significance level of 0.10 assuming an ABS coverage rate of 95 percent. Hybrid frames (such as the one investigated for NSDUH) would share an even larger proportion of cases with the FE frame because segments below the designated threshold would be field enumerated. Therefore, the hybrid FE and ABS frame would have even less coverage bias. For the hybrid frame investigated for NSDUH, the authors examined coverage bias among several subgroups, including rural areas and group quarters, and found no substantive differences in the estimates.

## **Development of a brief mental health impairment scale using a nationally representative sample in the USA**

**CITATION:** Novak, S. P., Colpe, L. J., Barker, P. R., & Gfroerer, J. C. (2010). Development of a brief mental health impairment scale using a nationally representative sample in the USA. *International Journal of Methods in Psychiatric Research*, 19 (Suppl. 1), 49-60. doi:10.1002/mpr.313

**PURPOSE/OVERVIEW:** A psychometric analysis was conducted to reduce the number of items needed to assess the disability associated with mental disorders using the World Health Organization Disability Assessment Schedule (WHODAS). The WHODAS was to be used in the Substance Abuse and Mental Health Services Administration's National Survey on Drug Use and Health (NSDUH), beginning in 2008, as part of a screening algorithm to produce estimates of the prevalence of serious mental illness (SMI) in the U.S. adult population. The goal of this paper was to create a parsimonious screening scale from the full 16-item WHODAS that was administered to 24,156 respondents aged 18 or older in the 2002 NSDUH.

**METHODS:** The authors used the 2002 NSDUH containing multiple measures of psychiatric symptoms, impairments, and mental health service use to include the following: (a) a series of disorder-specific items that had been determined to predict with certain probability meeting 12-month criteria for *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV), disorders from the World Health Organization (WHO) Composite International Diagnostic Interview Short Form (CIDI-SF) scales; (b) the Kessler-6/Kessler-10 (K6/K10) scales of nonspecific psychological distress; and (c) NSDUH's adult mental health service utilization module. These measures were used to identify and exclude members of the study sample who would likely not meet the criteria for a DSM disorder because associated mental health impairments would not apply to those respondents. The authors restricted the analyses to those respondents with the following characteristics: (1) endorsing a disorder-specific item from the

CIDI-SF screener, (2) endorsing "most" or "some of the time" for any one of the nonspecific distress items on the K6/K10 scale, (3) seeing a mental health professional in the past 12 months, (4) staying in a hospital or facility overnight for mental health issues in the past 12 months, (5) receiving outpatient treatment for mental health issues in the past 12 months, and (6) taking prescription medication for a mental condition in the past 12 months.

**RESULTS/CONCLUSIONS:** Exploratory factor analyses showed that WHODAS responses were unidimensional. A two-parameter polytomous Item Response Theory model showed that all of the 16 WHODAS items had good item discrimination (slopes greater than 1.0) for each response option. Analysis of item difficulties and differential item function across sociodemographic categories was then used to select a subset of 8 items to create a short version of the WHODAS. The Pearson correlation between scores in the original 16-item and reduced 8-item WHODAS scales was 0.97, documenting that the vast majority of variation in total scale scores was retained in the reduced scale.

### **The validity of State survey estimates of binge drinking**

**CITATION:** Paschall, M. J., Ringwalt, C. L., & Gitelman, A. M. (2010). The validity of State survey estimates of binge drinking. *American Journal of Preventive Medicine*, 39(2), 179-183. doi:10.1016/j.amepre.2010.03.018

**PURPOSE/OVERVIEW:** State survey-based estimates of binge drinking are useful to estimate the need for alcohol prevention and treatment services and to evaluate the effects of State alcohol control policies. However, because of declining survey response rates, there is growing concern about the validity of State survey estimates of binge drinking. This study examines the construct validity of State survey-based prevalence estimates of binge drinking.

**METHODS:** The authors used the State prevalence estimates of binge drinking in the past 30 days for 1999, 2001, 2003, 2005, and 2007 from published reports or public use data for the National Survey on Drug Use and Health, the Behavioral Risk Factor Surveillance System Survey, and the Youth Risk Behavior Survey. Construct validity was assessed in 2009 by examining correlations between these survey estimates and State per capita alcohol consumption levels (based on sales data for beer, wine, and spirits) and the percentage of drivers with a blood alcohol concentration of at least 0.08 who were in fatal motor vehicle crashes.

**RESULTS/CONCLUSIONS:** An estimated 88 percent of the correlations between State survey-based binge drinking estimates and per capita alcohol sales data were significant and moderate to strong ( $r \geq 0.30$ , range = 0.16 – 0.60). Similarly, 86 percent of the State survey binge drinking estimates were moderately or strongly correlated with the percentage of drivers in fatal crashes who had a blood alcohol concentration that was greater than or equal to 0.08 (range = 0.11 – 0.60). Based on the results, the authors concluded that State survey-based estimates of binge drinking have construct validity and therefore can be used to investigate relationships between State alcohol policies and other State characteristics and the prevalence of this behavior.

## The implications of geocoding error on address-based sampling

CITATION: Shook-Sa, B. E., McMichael, J. P., Ridenhour, J. L., & Iannacchione, V. G. (2010). The implications of geocoding error on address-based sampling. In *Proceedings of the 2010 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Vancouver, British Columbia, Canada* (pp. 3303-3312). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: In-person surveys that use address-based sampling (ABS) are often based on area segments defined by census geography rather than postal geography. Census geography enables more accurate inclusion of demographic information in the sample selection procedures and the use of frame supplementation methods to increase coverage. However, area frames based on census geography contain more frame error than frames based on postal geography because addresses must be allocated (i.e., geocoded) into area segments. When addresses are incorrectly geocoded into area segments, sampling inefficiencies occur.

METHODS: The authors examined data from the 2009 National Survey on Drug Use and Health to determine the extent of geocoding error in sampled segments and its implications on coverage and efficiency of area frame samples.

RESULTS/CONCLUSIONS: Geocoding accuracy at the segment level was quite poor and varied significantly by urbanicity. Rural segments had a much higher rate of undercoverage geocoding error (23.4 percent) compared with urban segments (7.5 percent). Geocoding accuracy improved significantly at the census block group level for both rural and urban segments, with 99.3 percent of addresses geocoding into the correct census block group (99.8 percent urban, 96.5 percent rural). These findings should be considered when designing ABS studies that are based on census geography. Geocoding error can be a significant source of undercoverage and sampling inefficiencies if segments are smaller than census block groups, especially in rural areas. Several characteristics of addresses and segments are related to segment-level geocoding error. Segment-level geocoding is more accurate in urban areas than in rural areas. Geocoding error also varies by census division, postal route and delivery type, the area of the segment, the proportion of new homes, and median home values within the segment.

*This page intentionally left blank*

# 2011

## **Errors in the recorded number of call attempts and their effect on nonresponse adjustments using callback models**

CITATION: Biemer, P., Chen, P., & Wang, K. (2011). Errors in the recorded number of call attempts and their effect on nonresponse adjustments using callback models. In *Proceedings of the International Statistics Institute, 58th World Statistical Congress, 2011, Dublin* (Session IPS033) (pp. 532-541). Cork, Ireland: Central Statistics Office.

PURPOSE/OVERVIEW: As the number of callbacks increases, the response propensity also increases. Therefore, the estimated response propensity for an individual case is a function of the number of callbacks. This paper considers the accuracy of callback data and the effect of errors in these data on the parameter estimates obtained from callback modeling.

METHODS: The callback model considered in this work was proposed by Biemer, Chen, and Wang (2010) for the 2006 National Survey on Drug Use and Health (NSDUH). The authors formed 20 response propensity strata from the response propensities obtained from the current NSDUH response propensity model, then they applied the callback model within each propensity stratum. Callback model estimates were obtained for each propensity stratum, then weighted together to obtain the callback model-adjusted estimate for the entire sample.

RESULTS/CONCLUSIONS: Level of effort (LOE) data are useful for nonresponse modeling if they are reasonably accurate. Callback response models assume that all callbacks to potential respondents are recorded and can be defined and captured in a standard way. If the LOE data are not recorded accurately, estimates of the response propensity from callback models will be biased. In turn, survey estimates incorporating these response propensity adjustments will also be biased.

## **Identifying common verbatim errors through use of field observations**

CITATION: Clark, C., McHenry, G., Williams, D., & Painter, D. (2011). Identifying common verbatim errors through use of field observations. In *Proceedings of the 2011 Joint Statistical Meetings, American Statistical Association, AAPOR 2011, Miami Beach, FL* (pp. 5965-5978). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: Training interviewers on the importance of reading all questions verbatim is one important way to ensure that standardized interviewing procedures are followed during each interview. Knowing which questions interviewers are most likely to misread allows project staff to improve training programs and reduce verbatim reading errors. This paper reviews the verbatim errors and exact questions observed not being read verbatim. The authors provide a summary of the types of questions with the most verbatim errors observed on the National Survey on Drug Use and Health (NSDUH) and offer possible explanations for these

errors. The types of questions with high verbatim errors observed on NSDUH may be indicative of common verbatim errors committed on other surveys as well.

**METHODS:** This study examined NSDUH field observation data collected across eight quarters of work, from January 2009 through December 2010, with 336 total observations made based on experience and results from previous observations. Verbatim errors were possible on 177 screening and interview screens. Observers noted which screens were not read verbatim during the observations and uploaded this information to a Web-based case management system (CMS). For all of 2009 and quarters 1 and 2 of 2010, items marked as verbatim errors were identified, and observer notes were manually reviewed to identify which screens specifically were not read verbatim. In quarter 3 of 2010, a new system was implemented to allow observers to select from a list those screens that were not read verbatim. For quarters 3 and 4 of 2010, all questions not read verbatim were aggregated from results entered by observers through the CMS. Data from both methods were aggregated for this study.

**RESULTS/CONCLUSIONS:** Overall, most screens were read verbatim most of the time. The total verbatim error rate for screenings was 2.9 percent, and the total verbatim error rate for interviews was 1.19 percent. Of the 13 screens observed not being read verbatim at least 5 percent of the time, 2 screens were from the iPAQ screening and 11 screens were from the CAPI portions of the interview. These 13 screens can be categorized into three screen types: instructional text, question not completed by interviewer, and question followed by text.

## **Redesigning contact materials for the National Survey on Drug Use and Health (NSDUH)**

**CITATION:** Currivan, D., Kennet, J., Painter, D., & Peytcheva, E. (2011). Redesigning contact materials for the National Survey on Drug Use and Health (NSDUH). In *Proceedings of the 2011 Joint Statistical Meetings, American Statistical Association, AAPOR 2011, Miami Beach, FL* (pp. 5783-5797). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** Designing survey contact materials presents a challenge for survey practitioners to determine the content and features most likely to encourage participation among recipients. For general household surveys, limited research exists to inform decisions on whether specific text and graphics in the contact materials are likely to be effective in facilitating cooperation and avoiding refusals. As a result, designing contact materials requires survey researchers to combine relevant data, knowledge, and experience to construct effective documents. This paper summarizes recent efforts to redesign the primary contact materials for the National Survey on Drug Use and Health (NSDUH).

**METHODS:** Three methods were used in redesigning the advance letter envelope, the advance letter, and the question and answer (Q&A) brochure for NSDUH. First, researchers developed alternative versions for each of these contact materials to address potential limitations of the current materials. Second, the current and alternative versions of the contact materials were submitted for expert review and feedback. Third, professional moderators conducted 17 focus groups with populations representing diverse demographics (such as gender, race, education, income, urbanity, country of origin for Spanish-language focus groups) across five metropolitan

areas—Chicago, Dallas-Fort Worth, Los Angeles, Raleigh-Durham, and Washington, DC. In order to ensure representation of the population who speak primarily English and those who speak primarily Spanish, 11 of the focus groups were conducted in English and 6 were conducted in Spanish to discuss participants' reactions to the different versions of the materials.

**RESULTS/CONCLUSIONS:** For the lead letter envelope, the majority of focus group participants preferred the larger envelope and stated that they would open the envelope because the "U.S. Department of Health and Human Services (DHHS)" logo led them to believe that the mailing was important. For the lead letter text, there was discrepancy in preference between the Spanish-speaking group and the English-speaking group participants. Based on these results, the authors recommended a hybrid version of the text that would combine the preferred text of each letter, while avoiding text considered to be problematic. Mixed preferences were shown for the signature and addressing of the letter (i.e., "Resident of \_\_\_\_ county"). For the lead letter graphics, focus group participants offered mixed preferences for which version they preferred. None of the versions garnered majority approval in either the English- or Spanish-speaking groups, and specific elements of the lead letter graphics seemed to heavily influence participant preferences. For the Q&A brochure, although a majority of the focus group participants preferred the alternative version, preferences differed between the English- and Spanish-speaking groups. Feedback on the Q&A brochures indicated that the alternative version had promise, but useful elements from the current version could be incorporated into the alternative version.

### **Estimating mental illness in an ongoing national survey**

**CITATION:** Gfroerer, J., Hedden, S., Barker, P., Bose, J., & Aldworth, J. (2011). Estimating mental illness in an ongoing national survey. In S. J. Blumberg & T. P. Johnson (Eds.), *Tenth Conference on Health Survey Research Methods: Conference proceedings* (pp. 35-41). Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

**PURPOSE/OVERVIEW:** The demand for more frequent and detailed data on mental illness increased with the passage of the 1992 Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) Reorganization Act. This legislation created the Substance Abuse and Mental Health Services Administration (SAMHSA) and required it to develop a definition and methodology for estimating serious mental illness (SMI) among adults, by State. In 2006, a technical advisory group recommended that the National Survey on Drug Use and Health (NSDUH) be modified to produce estimates of SMI among adults by supplementing the Kessler-6 (K6) psychological distress module with questions on functional impairment. The data from these short scales then would be used to estimate SMI using a statistical model based on clinical psychiatric interviews conducted on a subsample of NSDUH respondents. SAMHSA began methodological development and testing to implement these enhancements, referred to as the Mental Health Surveillance Study (MHSS), in NSDUH in 2008 (Colpe et al., 2010). This paper provides an overview of the NSDUH design; the development, implementation, and initial results of the MHSS; and plans for evaluating, improving, and utilizing the MHSS.

METHODS: N/A.

RESULTS/CONCLUSIONS: N/A

## **Identifying causes of verification refusals on a large nation-wide field study using a multilevel model**

CITATION: Williams, D., Touarti, C., Clark, C., & Butler, J. (2011). Identifying causes of verification refusals on a large nation-wide field study using a multilevel model. In *Proceedings of the 2011 Joint Statistical Meetings, American Statistical Association, AAPOR 2011, Miami Beach, FL* (pp. 5837-5848). Alexandria, VA: American Statistical Association.

PURPOSE/OVERVIEW: Accurate verification data are an essential part of field studies. In particular, verification offers assurance to survey sponsors and the public that data are valid and reliable. On the National Survey on Drug Use and Health (NSDUH), respondents are asked to provide contact information so that project staff may call to check on the quality of completed household screenings and interviews. Refusal of such information by respondents impedes the ability to verify fieldwork or, at best, introduces delays and added expense to the process. Identifying the causal factors for the absence of verification contact data allows for remedial actions, thus reducing costs and increasing quality. This paper presents the results of an analysis of verification refusals from the 2009 NSDUH.

METHODS: The authors used logistic regression to examine the effects of field interviewer (FI) performance measures and area demographic characteristics on the collection of verification contact information. FI performance was measured by production, cost, and data quality indicators. The demographic characteristics of sampling areas, or segments, were based on census data. The authors discussed the effect that each of these variables had on verification data in order to identify explanations for patterns in verification refusal rates.

RESULTS/CONCLUSIONS: Many explanations for high verification refusal rates are possible. However, explanations from the field are often subjective and inconclusive. This analysis explored objective measures of community characteristics and interviewer performance to explain the causes of verification refusals. Initial findings in the logistic regression indicated that a number of community demographics and interviewer characteristics, including interviewer performance and falsification, were significantly related to verification refusals. However, this analysis also revealed regional differences that need to be explored. Adjusting for census region showed that the percentage of group quarters units, percentage Hispanic, and the interviewer performance score had a noticeable impact on verification refusal rates. When examining an interviewer's verification refusal rate, the percentage of the segment population that is Hispanic and living in group quarters units, and the FI's performance in respect to response rates, data quality errors, and production costs, should all be taken into account. Furthermore, attention should be given to FIs with higher performance scores to understand their success in collecting verification information.

## **Abuse and dependence on prescription opioids in adults: A mixture categorical and dimensional approach to diagnostic classification**

CITATION: Wu, L. T., Woody, G. E., Yang, C., Pan, J. J., & Blazer, D. G. (2011). Abuse and dependence on prescription opioids in adults: A mixture categorical and dimensional approach to diagnostic classification. *Psychological Medicine*, 41(3), 653-664.  
doi:10.1017/s0033291710000954

PURPOSE/OVERVIEW: For the emerging *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition (DSM-V), it has been recommended that dimensional and categorical methods be used simultaneously in diagnostic classification; however, little is known about this combined approach for abuse and dependence.

METHODS: Using data ( $n = 37,708$ ) from the 2007 National Survey on Drug Use and Health (NSDUH), the authors examined *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV), criteria for prescription opioid abuse and dependence among nonprescribed opioid users ( $n = 3,037$ ) by using factor analysis (FA), latent class analysis (LCA, categorical), item response theory (IRT, dimensional), and factor mixture (hybrid) approaches.

RESULTS/CONCLUSIONS: A two-class factor mixture model (FMM) that combines features of categorical latent classes and dimensional IRT estimates empirically fitted more parsimoniously to abuse and dependence criteria data compared with models from FA, LCA, and IRT procedures, respectively. This mixture model included a severely affected group (7 percent) with a comparatively moderate to high probability (0.32 – 0.88) of endorsing all abuse and dependence criteria items and a less severely affected group (93 percent) with a low probability (0.003 – 0.16) of endorsing all criteria. The two empirically defined groups differed significantly in terms of the pattern of nonprescribed opioid use, comorbid major depression, and substance abuse treatment use. Furthermore, an FMM integrating categorical and dimensional features of classification fit better to DSM-IV criteria for prescription opioid abuse and dependence in adults than a categorical or dimensional approach. The authors suggested future research to examine the utility of this mixture classification for substance use disorders and treatment response.

*This page intentionally left blank*

## 2012

### **Innovative recruitment using online networks: Lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (webRDS) strategy**

CITATION: Bauermeister, J. A., Zimmerman, M. A., Johns, M. M., Glowacki, P., Stoddard, S., & Volz, E. (2012). Innovative recruitment using online networks: Lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (webRDS) strategy. *Journal of Studies on Alcohol and Drugs*, 73(5), 834-838.

PURPOSE/OVERVIEW: The authors used the Web version of a respondent-driven sampling (webRDS) strategy to recruit a sample of young adults aged 18 to 24 and examined whether this strategy would result in alcohol and other drug (AOD) prevalence estimates comparable with national estimates from the National Survey on Drug Use and Health (NSDUH).

METHODS: The authors recruited 22 initial participants (seeds) via Facebook to complete a Web survey examining AOD risk correlates. Sequential, incentivized recruitment continued until they achieved the desired sample size. After correcting for webRDS clustering effects, the authors contrasted participants' AOD prevalence estimates (past 30-day use) to 2009 NSDUH estimates by comparing the 95 percent confidence intervals of the prevalence estimates.

RESULTS/CONCLUSIONS: The authors found comparable AOD prevalence estimates between their study sample and NSDUH for the past 30-day use of alcohol, marijuana, cocaine, Ecstasy (3,4-methylenedioxymethamphetamine, or MDMA), and hallucinogens. Cigarette use was lower than found in NSDUH's estimates. The webRDS may be a suitable strategy to recruit young adults online. The authors discussed the unique strengths and challenges that may be encountered by public health researchers using webRDS methods.

### **Comparing and evaluating youth substance use estimates from the National Survey on Drug Use and Health and other surveys**

CITATION: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. (2012). *Comparing and evaluating youth substance use estimates from the National Survey on Drug Use and Health and other surveys* (HHS Publication No. SMA 12-4727, Methodology Series M-9). Rockville, MD: Substance Abuse and Mental Health Services Administration.

PURPOSE/OVERVIEW: This report describes the results of an effort to gain a better understanding of substance use data provided by adolescents on three Federal surveys by comparing National Survey on Drug Use and Health (NSDUH) estimates for adolescents with estimates from two other large-scale surveys: Monitoring the Future (MTF) and the Youth Risk Behavior Survey (YRBS), which is a component of the Youth Risk Behavior Surveillance System (YRBSS). The goals of this study were (1) to help users understand the reasons for

differences in estimates across the different surveys, (2) to facilitate accurate interpretation of results from these surveys; and (3) to improve the understanding of the true nature and extent of youth substance use in the United States.

**METHODS:** The authors examined estimates from NSDUH, a survey covering individuals aged 12 or older in the civilian, noninstitutionalized population of the United States, and two school-based surveys: the MTF, which surveys 8th, 10th, and 12th graders, and the YRBS, which covers the 9th through 12th grades. Where these surveys yielded different estimates of substance use, the report examined the possible reasons for these differences. Estimates for NSDUH and MTF were based on combined data from the respective 2002 through 2008 surveys. YRBS estimates were based on combined data from the 2003, 2005, and 2007 surveys. Thus, estimates in this report represented annual averages.

**RESULTS/CONCLUSIONS:** In most instances, NSDUH estimates of substance use for students in school were lower than corresponding estimates from the MTF and YRBS. YRBS estimates for 10th and 12th graders tended to be higher than MTF estimates. In general, these three surveys showed similar findings on which subgroups of adolescents had relatively higher or lower substance use estimates (e.g., in all three surveys, males in the 12th grade were more likely than females in this grade to be current users of cigarettes, marijuana, and cocaine). NSDUH and MTF, which are conducted annually, generally provided similar findings about changes over time (i.e., trends) in the prevalence of use of cigarettes, alcohol, and marijuana among 12th graders.

The reasons for discrepancy in the estimates can be explained by multiple reasons. First, conducting an interview in an adolescent's home environment has an inhibitory effect on adolescent substance users' willingness to report use. In contrast, youths could perceive that an interview in a classroom at school is more private than an interview at home. Additionally, peer presence in classroom settings could lead to some overreporting by youths in school-based surveys. Factors besides interview privacy also could contribute to lower estimates of adolescent substance use in NSDUH than in MTF or YRBS, and lower estimates in MTF than in YRBS. These other factors include the focus of the survey (e.g., primary focus on substance use or on broader health topics), how prominently substance use is mentioned when a survey is presented to parents and adolescents, procedures for obtaining parental permission for their children to be interviewed, assurances of confidentiality, the placement and context of substance use questions in the interview, the survey mode (e.g., computer-assisted interviewing with skip patterns or paper-and-pencil questionnaires), and the question structure and wording. Specifically, some NSDUH respondents may realize early during their interview that if they answer "no" to the initial filter questions about lifetime substance use, they can avoid having to answer subsequent questions and therefore will finish the interview in less time. The YRBS questionnaire does not have these kinds of skip patterns, and the MTF questionnaire uses skip patterns minimally. In addition, students taking a survey in a classroom administration setting may not be motivated to finish sooner if they otherwise have to stay until the end of the class period. Different designs also are likely to cause differing levels and patterns of coverage and nonresponse bias, which may or may not be alleviated by weighting adjustments or imputation procedures.

Although it is necessary to continue efforts to understand the impact of these factors, it is also important to recognize the critical contributions that each of the surveys makes in research and

policy development. The differences in the survey designs and procedures can be viewed as a strength because no single survey can adequately cover the full range of issues associated with substance use among adolescents.

## **National Survey on Drug Use and Health: Sample redesign issues and methodological studies**

CITATION: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. (2012, March). *National Survey on Drug Use and Health: Sample redesign issues and methodological studies* (RTI 0209009.486.001 and RTI 0211838.108.006.005, prepared by RTI International for the Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality, under Contract Nos. 283-2004-00022 and HHSS283200800004C). Rockville, MD: Substance Abuse and Mental Health Services Administration.

PURPOSE/OVERVIEW: The National Survey on Drug Use and Health (NSDUH), sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA), is a national survey of the U.S. civilian, noninstitutionalized population aged 12 or older. To continue producing timely and relevant data, SAMHSA's Center for Behavioral Health Statistics and Quality (CBHSQ) must update NSDUH periodically to reflect changing substance use and mental health behavior and data needs. In 2012, CBHSQ was planning to implement a redesigned NSDUH sample beginning with the 2014 survey and a redesigned questionnaire beginning with the 2015 survey. In preparation for the redesign, the authors examined the feasibility and impacts of a variety of sample design changes on survey costs and data precision. In this report, the authors examined the following issues: (1) optimal cluster size, (2) optimal sample distribution with respect to demographic and geographic groups, (3) pros and cons of collecting data on a more continuous basis (with little or no downtime within data collection periods), (4) feasibility of conducting the survey every other year, (5) implications of expanding the target population to include children under the age of 12, and (6) feasibility of a flexible design.

METHODS: N/A.

RESULTS/CONCLUSIONS: The results and conclusions for each of the six issues are as follows:

1. Regarding the optimal cluster size, an analysis of the cost-variance trade-offs suggested that a larger sample per segment than what was currently being used for NSDUH would be nearly optimal for producing national estimates. In the formal optimization, which solved across several drug use, dependence, and treatment variables for multiple age groups simultaneously, it was determined that the optimal cluster size was around 29 individuals per segment. Univariate calculations confirmed that this cluster size would provide as nearly as precise estimates at a greatly reduced cost (approximate savings of \$4 million).
2. Regarding the optimal sample distribution, a national design with a sample of 40,000 offered the greatest cost reductions, in large part because of the reduction in field staff required.

Hybrid designs still produced major cost reductions while preserving minimal sample sizes for some State estimates. On the other hand, a precision analysis favored a 50-State design with a smaller cluster size. State and substate estimates were not practical under the national design, and under the hybrid design, State estimates required pooling data over more years than were currently required. Unweighted overall response rates were expected to decrease for the age allocations that shifted the sample distribution toward older age groups.

3. Regarding the continuous data collection design, data collection on a more continuous basis would more accurately portray annual drug use if peaks and troughs in drug use occurred throughout the year; however, little evidence was available to suggest seasonal differences at the national level, although small differences in drug use may occur among subgroups throughout the year. A continuous data collection design would have a potential negative effect on response rates, training and field operation costs, staffing, and data quality, and it would have a potential positive effect on survey planning, such as implementing questionnaire changes, responding to major events, and achieving sample targets. A more continuous design would have minimal impacts on analysis and reporting unless case assignments were extended into the following data collection period and the following survey year.

4. Regarding a biennial survey, switching NSDUH from the current annual data collection and analysis to a biennial design would result in large cost savings. Although the average cost per interview would be expected to increase under a biennial design because of increased administrative and training activities and a "learning curve" for new and returning staff, it still would be less expensive to conduct the survey only every other year. However, a biennial design could affect the quality and timeliness of NSDUH data. The inability to retain field staff could contribute to an inconsistency in interviewer experience levels, which in turn could affect response rates and respondent reporting of sensitive items. Additionally, NSDUH would no longer have the ability to quickly respond to substantive issues or to conduct special analyses on the impact of national events; moreover, estimates produced using combined years of data, such as those for smaller geographic regions (e.g., States and substate areas) or for rare subpopulations (e.g., pregnant women), would no longer be timely and in some cases would become impractical.

5. Regarding expanding the target population, the main benefit of expanding the sample by including children under age 12 is the ability to provide accurate estimates for this age group. Under the current design, the undercoverage of past year initiates aged 11 or younger also affects the mean age at first use estimate. However, there are several issues to consider such as modification of screening software, increase of refusal rates, concern with data quality, questionnaire appropriate for the reading level, and IRB concerns.

6. Regarding a flexible design, from a sample design perspective, several options are possible, including hybrid designs, changing cluster sizes, address-based sampling (ABS), continuous or biennial data collection, and expanding the target population. The authors indicated that a decision on this design option should be made by considering the impacts of the survey design, such as loss of data comparability (i.e., trends), decreased precision, and changes in operational procedures that may have cost implications.

## **Imputation using the other pair member**

**CITATION:** Frechtel, P., Scott, V., Couzens, A., Moore, A., & Bose, J. (2012). Imputation using the other pair member. In *Proceedings of the 2012 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, San Diego, CA* (pp. 4248-4258). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** In the National Survey on Drug Use and Health (NSDUH), zero, one, or two people are selected from each selected household. "Pairs" account for about 60 percent of the annual sample, over 90 percent of whom are members of the same family. The responses to some NSDUH questions have high positive correlations between pair members, especially the questions about family-level characteristics. The current imputation method, predictive mean neighborhood (PMN), does not exploit this correlation, even when the value for one pair member is missing and the value for the other pair member is not missing.

**METHODS:** The authors investigated (1) a method for identifying variables for which the other pair member's response may be a better choice for imputation than the PMN-assigned value; and (2) for these variables, a method for identifying exact conditions under which the other pair member's response may be used.

**RESULTS/CONCLUSIONS:** A crude assessment of feasibility applied to all variables that undergo imputation suggested that the income and health insurance variables were the best candidates for this method. Most of these variables had nontrivial numbers of missing values, and the other-pair-member method (i.e., item nonrespondents paired with item respondents) applied to adequate numbers of these cases. The proportion of pairs whose responses agreed after PMN imputation was usually considerably lower than the proportion of responding pairs whose responses agreed, suggesting that the other-pair-member method was preferable. For the income and health insurance variables, two factors were considered: (1) the type of pair (definitely in the same family or not) and (2) the number of days between the presentation of the questions to the pair members. After considering these factors and using logistic regression models, the authors concluded that for the nine income variables that stored data at the level of the family in the household, the use of the other pair member's value in imputation appeared to improve the quality of imputed data as long as the pair members were in the same family. For the eight health insurance variables that underwent imputation, the use of the other pair member's value in imputation did not appear to improve data quality because the deterministic nature of this method was inappropriate given the rate of disagreement between responding pair members.

## **Methodological considerations in estimating adolescent substance use**

**CITATION:** Gfroerer, J., Bose, J., Kroutil, L., Lopez, M., & Kann, L. (2012). Methodological considerations in estimating adolescent substance use. In *Proceedings of the 2012 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, San Diego, CA* (pp. 4127-4140). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** The Federal Government relies primarily on three major national surveys in tracking adolescents' use of alcohol, tobacco, and illicit drugs. The National Survey on

Drug Use and Health (NSDUH), the Monitoring the Future (MTF) study, and the national Youth Risk Behavior Survey (YRBS), a component of the Youth Risk Behavior Surveillance System (YRBSS), showed similar trends in substance use between 2002 and 2012, but the surveys showed significant differences in the levels of use of some substances. The school-based surveys (MTF and YRBS) generally reported higher rates of use than the household survey (NSDUH). Prior methodological research explored how various design features such as mode, setting, privacy, question wording, and coverage may affect substance use estimates from the surveys. This paper explores these factors based on new analyses of combined 2002-2008 data from the surveys.

**METHODS:** The authors first examined estimates of substance use from NSDUH with estimates from the MTF and YRBS. Differences in levels of use, demographic patterns of use, and trends in use were examined. These comparisons were made within specific grades of school. Secondly, they used NSDUH data to examine the potential effects on substance use estimates because of school dropouts and youths who were absent from school. Then they examined relationships between the privacy of NSDUH interviews and substance use estimates. NSDUH and MTF estimates were based on combined data from their respective 2002 through 2008 surveys. YRBS estimates were based on combined data from its 2003, 2005, and 2007 surveys.

**RESULTS/CONCLUSIONS:** The differences in survey designs across the three surveys have had an impact on the surveys' estimates. For example, conducting an interview in an adolescent's home environment could inhibit adolescent substance users' willingness to report use because of their parents' presence. In contrast, youths could perceive an interview in a classroom at school to be more private than an interview at home. Another possible explanation that warrants further study is whether the presence of friends in the classroom setting also leads to some overreporting by youths. Additional factors could also affect reporting levels, such as the focus of the survey (e.g., primary focus on substance use or on broader health topics), how prominently substance use is mentioned when a survey is presented to parents and adolescents, assurances of confidentiality, the placement and context of substance use questions in the interview, the survey mode (e.g., computer-assisted interviewing with skip patterns or paper-and-pencil questionnaires), and the question structure and wording.

## **CBHSQ Data Review: Comparison of NSDUH mental health data and methods with other data sources**

**CITATION:** Hedden, S., Gfroerer, J., Barker, P., Smith, S., Pemberton, M. R., Saavedra, L. M., Forman-Hoffman, V. L., Ringeisen, H., & Novak, S. P. (2012, March). *CBHSQ Data Review: Comparison of NSDUH mental health data and methods with other data sources*. Rockville, MD: Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality.

**PURPOSE/OVERVIEW:** This report compares adult mental health prevalence estimates generated from the 2009 National Survey on Drug Use and Health (NSDUH) with estimates of similar measures generated from other national data sources. It also describes the methodologies of the different data sources and discusses the differences in survey design and estimation that may contribute to differences among these estimates.

**METHODS:** The comparison data sources included the 2001 to 2003 National Comorbidity Survey Replication (NCS-R), 2001 to 2002 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), 2007 Behavioral Risk Factor Surveillance System (BRFSS), 2008 National Health Interview Survey (NHIS), 2008 Medical Expenditure Panel Survey (MEPS), and 2008 Uniform Reporting System (URS). The authors discussed mental disorder indicators such as past year (12 months preceding survey interview) serious mental illness (SMI), past year any mental illness (AMI), past month (30 days preceding survey interview) serious psychological distress (SPD), past year major depressive episode (MDE), and past year suicidality (suicidal thoughts and behaviors).

**RESULTS/CONCLUSIONS:** The authors concluded that substantial methodological differences across the data sources with unmeasured effects on estimation make it difficult to determine which of the various estimates is "best." Agreement between the data sources is not expected, and the discrepancy does not reduce the importance of these studies in providing a comprehensive picture of mental health in the United States as each study was designed for a different purpose. For example, NSDUH not only collects information on mental health, but also is the primary source of data collected on substance use and substance use disorders. NSDUH also includes extensive data on demographics, physical health, receipt of treatment for mental disorders and substance use, and various other topics relevant to mental health. Therefore, NSDUH data have been used to examine the association between mental health issues and a variety of correlates. Furthermore, the large sample size and the annual nature of NSDUH allow for precise and up-to-date estimates of mental health indicators for various subpopulations (e.g., specific racial/ethnic groups) and the capability of tracking trends over time. BRFSS, NHIS, and MEPS also are designed so that trends in estimates can be produced. In addition, NSDUH and BRFSS allow for State-specific estimates of mental disorders. The authors emphasized that, when comparing survey estimates of mental health indicators across multiple data sources, one should consider the methodological differences. Understanding the differences in methodology, survey mode, and specific measures used to assess different mental health indicators across these surveys can help to provide context for understanding and interpreting the various prevalence estimates that have been published from these surveys.

## **Evaluation of small area estimates of substance use in the National Survey on Drug Use and Health**

**CITATION:** Hughes, A., Vaish, A., Sathe, N., & Spagnola, K. (2012). Evaluation of small area estimates of substance use in the National Survey on Drug Use and Health. In *Proceedings of the 2012 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, San Diego, CA* (pp. 4661-4674). Alexandria, VA: American Statistical Association.

**PURPOSE/OVERVIEW:** Small area estimation (SAE) methods are used to produce State and substate estimates of substance use and mental disorders using data from a major U.S. behavioral health survey. State and local policymakers use these estimates to understand the nature and extent of the problem and to justify funding for substance use prevention and treatment programs in their jurisdictions. Design-based estimates could be used as an alternative; they are less expensive than small area estimates and take less time to produce. Thus, it is important to

determine how small area estimates compare with their design-based counterparts in terms of accuracy and precision.

**METHODS:** For this validation study, pooled 2009 and 2010 National Survey on Drug Use and Health (NSDUH) data were used. Each of the eight large States (i.e., California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas) has 48 State sampling regions (SSRs) or strata. The authors created various pseudo small States within each of the eight large States by pooling 2 years of NSDUH data so that the pseudo small State sample sizes ranged from 75 to 300 for the 12 or older age group (or 25 to 100 per age group). Then they fit the current SAE models, produced small area estimates for each of the pseudo small States, and compared them with the corresponding large State design-based estimates. Each of the eight large States' design-based estimates were based on about 7,200 respondents for the 12 or older age group (about 2,400 respondents in each of the three age groups) and were considered as "true values" for comparison purposes.

**RESULTS/CONCLUSIONS:** This validation study attempted to measure the accuracy of small area estimates with respect to various small sample sizes and provide useful information about the quality of NSDUH's substate estimates. The authors noted that using the suppression rule on the age group-specific substate estimates, where sample sizes could be as low as 35, would ensure that only good quality estimates are published. A similar validation study was recommended in order to judge the quality of the substate year-to-year change estimates with respect to various small sample sizes in the future. The authors concluded that the use of SAE techniques to produce State and substate small area estimates provides substantial improvement in accuracy relative to their design-based counterparts.

## **A proposed hybrid sampling frame for the National Survey on Drug Use and Health**

**CITATION:** Iannacchione, V., McMichael, J., Shook-Sa, B., & Morton, K. (2012). *A proposed hybrid sampling frame for the National Survey on Drug Use and Health* (RTI 0209009, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-2004-00022). Research Triangle Park, NC: RTI International.

**PURPOSE/OVERVIEW:** The target population for the National Survey on Drug Use and Health (NSDUH) is the civilian, noninstitutionalized population of the United States aged 12 or older. The NSDUH sampling frame relies on field enumeration (FE) to identify dwelling units (DUs) that are eligible for selection and the half-open interval (HOI) frame-linking procedure to cover DUs missed during FE. Because of the rising costs associated with in-person surveys of the U.S. general population, address-based sampling (ABS) is increasingly viewed as a potential alternative. However, the cost savings afforded by the use of ABS over FE for NSDUH are tempered by research indicating that FE provides more complete coverage than ABS, especially in rural areas. This report summarizes research conducted by RTI to evaluate the coverage, cost, and implementation of the proposed hybrid sampling frame for NSDUH.

**METHODS:** RTI developed and piloted a hybrid sampling frame for NSDUH that uses ABS supplemented with the Check for Housing Units Missed (CHUM) frame-linking procedure in

area segments with adequate ABS coverage, but retains FE in segments where poor ABS coverage is anticipated. The authors investigated the trade-offs between coverage and cost savings as area segments are shifted from FE to ABS. The data used in the analyses came from the Mailing List Field Studies (MLFS) I and II, subsequent work developing and testing the CHUM procedure, and exploratory analyses on coverage prediction, group quarters (GQs) coverage, geocoding error, and potential supplemental sources of addresses.

**RESULTS/CONCLUSIONS:** Theoretically, the hybrid frame provides 100 percent coverage of the target population. In FE segments, the coverage is equivalent to the current NSDUH coverage rate. In ABS segments, DUs not included on the ABS frame are covered by the CHUM procedure. The only known sources of undercoverage occur when field staff incorrectly implement the CHUM and/or HOI procedures. ABS coverage of GQs is problematic. Therefore, segments with high concentrations of GQs should be allocated to FE. Geocoding error occurs when the geographic coordinates assigned to a DU do not correspond to its actual location. The cost savings afforded by the hybrid frame depend on how many segments are assigned to ABS. In general, the lower the ABS coverage threshold, the more segments will be allocated to ABS and the higher the cost savings. With proper training and monitoring, the hybrid sampling frame can be implemented in a way that reduces survey costs while maintaining NSDUH's high scientific standards. Further efficiencies can be gained by developing techniques that accurately allocate segments with low ABS coverage (e.g., segments with high concentrations of GQs) to the FE frame and by continuing to explore sources of supplemental addresses.

### **Adult current smoking: Differences in definitions and prevalence estimates—NHIS and NSDUH, 2008**

**CITATION:** Ryan, H., Trosclair, A., & Gfroerer, J. (2012). Adult current smoking: Differences in definitions and prevalence estimates—NHIS and NSDUH, 2008. *Journal of Environmental and Public Health*, 2012, 918368 [article number].

**PURPOSE/OVERVIEW:** This paper compares prevalence estimates and assesses issues related to the measurement of adult cigarette smoking between the National Health Interview Survey (NHIS) and the National Survey on Drug Use and Health (NSDUH).

**METHODS:** The authors compared data from the 2008 NHIS and the 2008 NSDUH on current cigarette smoking and current daily cigarette smoking among adults aged 18 years or older. They used the standard NHIS current smoking definition, which screens for lifetime smoking of 100 or more cigarettes. For NSDUH, both the standard current smoking definition, which does not screen, and a modified definition applying the NHIS current smoking definition (i.e., with screen) were used.

**RESULTS/CONCLUSIONS:** NSDUH consistently yielded higher current cigarette smoking estimates than NHIS and lower daily smoking estimates. However, with the use of the modified NSDUH current smoking definition, a notable number of subpopulation estimates became comparable between the two surveys. Younger adults and racial/ethnic minorities were most affected by the lifetime smoking screen, with Hispanics being the most sensitive to differences in smoking variable definitions among all subgroups. Differences in current cigarette smoking

definitions appeared to have a greater impact on smoking estimates in some subpopulations than others. Survey mode differences may also have limited intersurvey comparisons and trend analyses.

# 2013

## **Incorporating level of effort paradata in the NSDUH nonresponse adjustment process**

CITATION: Biemer, P., Chen, P., & Wang, K. (2013, June). *Incorporating level of effort paradata in the NSDUH nonresponse adjustment process* (RTI 0211838.108.106.010, prepared for the Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality, under Contract No. HHSS283200800004C). Research Triangle Park, NC: RTI International.

PURPOSE/OVERVIEW: The Substance Abuse and Mental Health Services Administration's planned changes for the National Survey on Drug Use and Health (NSDUH) in 2014 and 2015 provide an opportunity to investigate new approaches to nonresponse adjustment that may have the potential of improving overall NSDUH data quality in the face of declining response rates for both NSDUH and surveys in general. Compared with traditional methods for estimating the response propensity requiring information that is available for both respondents and nonrespondents, callback modeling can use variables that are known only for respondents. This report describes the results of an investigation of the callback modeling approach for adjusting for nonresponse.

METHODS: The methods used were adapted from Biemer and Link (2007) for NSDUH. The NSDUH interview process consists of a household screener used to enumerate household members and identify eligible respondents, followed by the selection of up to two members of the household for an interview. The focus of this report was on the main interview, and the analysis used the record of calls (ROC) data entered by interviewers for each visit to a sampled dwelling unit. Interviewers enter case status information using a handheld computer. Collected data elements include the interim or final outcome of the call (e.g., noncontact, refusal, completed screening, completed interview) and the time and day of the call attempt, which were automatically recorded by the handheld computer. All sample cases are ultimately categorized into a number of final case dispositions. In this report, these dispositions were collapsed into (1) interviewed, (2) final refusal, (3) final nonresponse other than refusal (e.g., language barriers, physically or mentally unable, or unavailable), or (4) censored.

RESULTS/CONCLUSIONS: The callback model estimates considered in this report showed little or no improvement over the generalized exponential model (GEM) and the GEM+ model estimates. The level of error in the callback data for callback modeling purposes was quite high and had the potential to seriously bias the results of the callback modeling approach to nonresponse adjustment. The limited simulation study confirmed that even a small degree of underreporting (e.g., 5 percent) was enough to appreciably bias the callback model estimates. Additionally, introduced by variations among supervisors, interviewers, and types of units in the levels of underreporting, the situation became somewhat intractable to address through model enhancements alone. The preferred solution would be to improve the quality of the callback data for modeling purposes at its source. However, changing the field procedures currently in use for

collecting these data and introducing additional quality checks to ensure accurate reporting of callbacks suitable for callback modeling would introduce additional burdens on interviewers.

### **Using level-of-effort paradata in non-response adjustments with application to field surveys**

CITATION: Biemer, P. P., Chen, P., & Wang, K. (2013). Using level-of-effort paradata in non-response adjustments with application to field surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1), 147-168. doi:10.1111/j.1467-985X.2012.01058.x

PURPOSE/OVERVIEW: In this paper, the authors consider the use of level-of-effort (LOE) paradata to model the nonresponse mechanism in surveys and to adjust for nonresponse bias, particularly bias that is missing not at random or nonignorable.

METHODS: The authors used an approach based on an unconditional maximum likelihood estimation (callback) model that adapts and extends the prior work to handle the complexities that are encountered for large-scale field surveys. A test of the "missingness at random" assumption was also proposed that can be applied to essentially any survey when LOE data are available. The nonresponse adjustment and the test for missingness at random were then applied and evaluated for a large-scale field survey—the National Survey on Drug Use and Health (NSDUH).

RESULTS/CONCLUSIONS: Although evidence on nonignorable nonresponse bias was found for NSDUH, the callback model could not remove it. One likely explanation of this result was error in the LOE data. This possibility was explored and supported by a field investigation and simulation study informed by data obtained on LOE errors.

### **Getting the most out of a limited sample size field test: Experiences from the National Survey on Drug Use and Health**

CITATION: Bose, J., Painter, D., Currivan, D., Kroutil, L., & Wang, K. (2013, November). Getting the most out of a limited sample size field test: Experiences from the National Survey on Drug Use and Health. In *Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) research conference*. Washington, DC: Federal Committee on Statistical Methodology.

PURPOSE/OVERVIEW: In this paper, the authors used experiences from the survey redesign process of the National Survey on Drug Use and Health (NSDUH), including a field test. They described ways in which even a limited-sample field test can inform decision making regarding the redesign or changing of a survey. The 2012 NSDUH field test was meant to test the revisions to the questionnaire and protocols. Additionally, the field test provided the first opportunity to see how the new items performed that had been added to the questionnaire in 2013. The new sample design allowed for continued national, State, and substate-level estimation comparable with estimation procedures from previous surveys while achieving cost savings. The primary change to the questionnaire was to update the prescription drug questions. Changes to the prescription drug questions included a new question structure that focuses on any use and misuse

(also known as nonmedical use) of specific prescription drugs in the past 12 months, a revised definition of misuse that includes overuse of prescribed medication, and a focus on currently available prescription drugs. Other planned changes to the questionnaire included a revised health module that contains new questions about medical conditions, a separate methamphetamine module, new items on sexual orientation, military families and disability, and other updates. Other changes to the survey included new contact materials and equipment, a change in the sponsor presented to the respondent (U.S. Public Health Service to the U.S. Department of Health and Human Services), use of on-screen pictures of prescription drugs and calendars instead of physical showcards, and moving some items from computer-assisted personal interviewing (CAPI) to audio computer-assisted self-interviewing (ACASI). These changes sought to achieve three main goals: (1) revise the questionnaire to address changing policy and research data needs, (2) modify the survey methodology to improve the quality of estimates and the efficiency of data collection and processing, and (3) where appropriate, maintain trends in core substance use estimates across survey years.

**METHODS:** Because of the limited sample size, it was difficult to isolate the impact of specific changes on the estimates. To overcome this difficulty and help make decisions regarding the 2015 survey design, the different sources of information were used as described below.

*Using NSDUH data that had already been collected:* Because the NSDUH 2012 field test sample was not large enough to accommodate one or more split samples, RTI used the corresponding main survey data that had been collected in two different time periods as comparison datasets: main survey data from all four quarters of 2011 and main survey data that had been collected during a period in 2012 that was similar to the data collection period for the field test. The 2011 data were used as a comparison group because the large sample size and data collection over an entire year would improve the precision of estimates and help identify any seasonality in the estimates. The 2012 main survey data were used as a comparison dataset because the data were collected during roughly the same time period as the field test and therefore would not be subject to annual changes in prevalence.

*Comparison with external data sources:* Items that were collected for the first time in the field test had no point of reference from the current NSDUH for comparison purposes. Therefore, these variables (e.g., height and weight, cellular phone coverage) were compared with the estimates in other large-scale Federal surveys, such as the National Health Interview Survey (NHIS) and the National Health and Nutrition Examination Survey (NHANES), including respondent self-reports and physical measures.

*Using timing data:* As part of the field test, timing data were also collected to get the estimates of the time taken to complete the entire survey and the individual module according to the respondent demographics (i.e., education attainment, age). These timing data also were available for evaluation if the survey was taking more or less time compared with the existing version.

*Using interviewer debriefing items:* A set of interviewer debriefing items were added that the interviewers had to complete after the data were collected but before they could close out a case. Examples included the following: "Did the respondent remember receiving the lead letter?" "What comments, if any, did the respondent make about the lead letter or in response to the lead letter?" "When did you give the respondent (or parent/guardian of youth respondent) the Q&A

[question and answer] brochure?" "What comments, if any, did the respondent (or parent/guardian) make about the Q&A [question and answer] brochure?"

*Using existing external research:* As part of attempts to decrease interviewer burden and increase respondent confidentiality, some items related to employment, health insurance, income, and education were moved from CAPI to ACASI administration, and notable differences were found in the estimates of sources of income and in health insurance coverage between the NSDUH field test and both of the NSDUH comparison datasets. Moving these items to ACASI in 2015 could be justified if the estimates based on ACASI data from the field test had either remained comparable with those from the main survey or if it could be demonstrated that ACASI administration improved the estimates. Improvements in the estimates based on ACASI administration could be demonstrated in one of two ways: (1) by showing that there was a theoretical basis for expecting these types of estimates to improve when the questions are self-administered; or (2) by comparing them with a gold-standard estimate. Literature review and comparison with the estimates in the other surveys (NHIS, the American Community Survey, and the Current Population Survey) were conducted.

*Examining rates of missing data:* Item nonresponse (i.e., missing data) can occur if (1) respondents do not know an answer (e.g., they do not remember how old they were when they first used a substance); or (2) they refuse to answer the question. NSDUH's rates of item nonresponse typically are fairly low. Therefore, any change in item nonresponse or having higher than expected levels of item nonresponse in new items may indicate that (1) there is a programming issue in the computer-assisted interviewing (CAI) instrument; (2) question wording or other issues with the question are adversely affecting the cognitive tasks associated with answering the question; or (3) the question wording or subject matter are too "loaded" or sensitive for respondents.

*Using prior NSDUH methodological work and "other specify" responses:* Existing methodological research also was used to make decisions. For example, the decision on the imputation methods to use for the field test were driven by the small sample size (and, therefore, a paucity of donors) and time. However, this decision could be made with relative confidence based on an earlier study that compared NSDUH estimates based on different imputation methods. "Other, specify" were also examined data to see whether respondents were selecting the "other category" at rates that were higher than expected and whether the "other specify" responses included those that were already available in the precoded response options.

**RESULTS/CONCLUSIONS:** The relatively small sample size for the field test affected the confidence in drawing certain conclusions from the data. For example, a lack of statistically significant differences between estimates in the field test and comparison datasets from the main survey did not provide certainty that there will be no effect in 2015 on NSDUH's substance use estimates, especially its trend data (e.g., estimates for alcohol, tobacco, marijuana, and cocaine). For certain individual prescription drugs or groups of prescription drugs with a common active ingredient, low estimates of misuse in the field test (or no respondents who reported misuse) may have been indicative of the small sample size rather than the actual prevalence in the population. The available sample size from the field test also may not allow the effects of changes to the prescription drug module on data in subsequent sections of the questionnaire that draw upon respondents' prior answers to the prescription drug items (e.g., substance dependence and abuse

and substance use treatment) to be adequately gauged. Additionally, because of the multiple changes in the field test (e.g., new contact materials, new equipment, changes to the questionnaire), firm conclusions about the most important reason for an observed difference, or the relative contributions of various changes to a result, cannot be drawn. In particular, one of the challenges in effectively using all of the data that were collected from the field test was making sure that there was sufficient time to process, analyze, and draw conclusions from the data.

### **Statistics on cannabis users skew perceptions of cannabis use**

CITATION: Burns, R. M., Caulkins, J. P., Everingham, S. S., & Kilmer, B. (2013). Statistics on cannabis users skew perceptions of cannabis use. *Frontiers in Psychiatry, 4*, 138. doi:10.3389/fpsy.2013.00138

PURPOSE/OVERVIEW: Collecting information about the prevalence of cannabis use is necessary, but not sufficient for understanding the size, dynamics, and outcomes associated with cannabis markets.

METHODS: This paper uses data from the National Survey on Drug Use and Health (NSDUH) conducted between 2002 and 2011 and the European Union Drugs Markets II (EUMII) Web survey conducted in 2012 about cannabis consumption in the United States and Europe to highlight (1) differences in inferences about subpopulations based on the measure used to quantify cannabis-related activity, (2) how different measures of cannabis-related activity can be used to more accurately describe trends in cannabis usage over time, and (3) the correlation between frequency of use in the past month and average grams consumed per use day.

RESULTS/CONCLUSIONS: Focusing on days of use instead of prevalence in the 2011 NSDUH showed substantially greater increases in U.S. cannabis use in recent years; however, the recent increase was mostly among adults, not youths. Relatively more rapid growth in use days also occurred among college-educated individuals and Hispanics. Additionally, data from a survey conducted in seven European countries based on the 2012 EUMII showed a strong positive correlation between frequency of use and quantity consumed per day of use, suggesting consumption was even more skewed toward the minority of heavy users than was suggested by days-of-use calculations.

### **Revised estimates of mental illness from the National Survey on Drug Use and Health**

CITATION: Center for Behavioral Health Statistics and Quality. (2013, November 19). Revised estimates of mental illness from the National Survey on Drug Use and Health. *The NSDUH Report*.<sup>2</sup>

PURPOSE/OVERVIEW: The Substance Abuse and Mental Health Services Administration (SAMHSA) has produced national, State, and substate estimates of mental illness among adults using its National Survey on Drug Use and Health (NSDUH) since 2008. These estimates are

---

<sup>2</sup> Also see the Kott et al. (2013) entry on a subsequent page of this report section.

based on the Mental Health Surveillance Study (MHSS), initiated in 2008. The overarching goal of the MHSS was to provide accurate estimates of serious mental illness (SMI) among adults aged 18 or older at both the national and State levels, with a secondary goal of providing accurate estimates of any mental illness (AMI). This report discusses recent improvements to the methods used to produce these estimates and provides revised 2011 estimates.

**METHODS:** The clinical diagnostic interview to assess the presence of mental disorders and functional impairment was administered to a subsample of NSDUH adult respondents within 2 to 4 weeks of the NSDUH interview. Each participant was assessed by a trained clinical interviewer over the telephone. A prediction model of SMI was developed from the clinical subsample of respondents using the short scales of psychological distress and functional impairment in combination with the clinical diagnostic data. Then the model was used to classify each of the NSDUH adult respondents as having SMI or AMI based on their responses to the distress and impairment scales. In the initial model using the 2008 MHSS subsample of approximately 750 respondents, the psychological distress and functional impairment indicators were used as predictor variables of SMI diagnosis. The model was then used to classify each of the NSDUH adult respondents as having SMI or AMI based on their responses to the distress and impairment scales. In the improved model developed in 2012, about 5,000 accumulated completed clinical assessments were used. This larger clinical sample enabled SAMHSA to evaluate the current methods for estimating mental illness to see if they could be improved. The evaluation of the methods included an assessment of the 2008 statistical model and the weights applied to the clinical interview sample used to develop the model.

**RESULTS/CONCLUSIONS:** Using combined 2008 to 2012 data, SAMHSA produced estimates using the initial method and the improved method and compared these estimates with those from the clinical interview sample. This comparison showed that estimates based on the improved methods were more similar to the clinical data than estimates based on the 2008 methods. Furthermore, estimates of SMI and AMI by age groups produced using the improved methods were closer to the estimates produced using the clinical data. Thus, revising the weights and adding new predictors to the model produced more accurate estimations of SMI and AMI.

### **Estimating substance abuse treatment: A comparison of data from a household survey, a facility survey, and an administrative data set**

**CITATION:** Gfroerer, J., Bose, J., Trunzo, D., Strashny, A., Batts, K., & Pemberton, M. (2013, November). *Estimating substance abuse treatment: A comparison of data from a household survey, a facility survey, and an administrative data set*. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC.

**PURPOSE/OVERVIEW:** This paper evaluates the coverage, overlap, biases, strengths, and weaknesses of three sources of data on the receipt of substance use treatment: the National Survey on Drug Use and Health (NSDUH), the National Survey of Substance Abuse Treatment Services (N-SSATS), and the Treatment Episode Data Set (TEDS). Each of these data systems collects data related to some of these key measures. The objective of this paper was to improve the understanding of the information on substance use treatment from these three data sources to inform future reporting of results and uses of the data. NSDUH, N-SSATS, and TEDS differ in

their intended goals and in their methods of data collection. A clear understanding of the differences between these three data systems, as well as their strengths and limitations, is necessary in order to maximize the usefulness of the systems and ensure the accurate interpretation of findings.

**METHODS:** The principal measures of treatment produced by the three data systems are not directly comparable. NSDUH estimates are generally presented in terms of individuals receiving treatment, N-SSATS data primarily describe the clients in treatment on a single reference day, and TEDS estimates are typically presented as admissions or discharges. However, using some of the additional items captured by each system and applying certain assumptions, measures can be constructed for comparison across the three data systems. The authors chose three such measures to focus on for this study: (1) number of clients in treatment on a single day (available in all three systems), (2) number of individuals receiving treatment within a 12-month period (NSDUH and TEDS), and (3) number of admissions over a 12-month period (N-SSATS and TEDS). Comparisons were made at the national and State levels, and reasons for similarities and differences were discussed.

**RESULTS/CONCLUSIONS:** Across the three measures of interest, TEDS data showed undercoverage relative to the N-SSATS and NSDUH estimates. The TEDS estimate of the single-day count was 46 percent of the N-SSATS count and 37 to 44 percent of the NSDUH estimates only. The TEDS estimate of individuals receiving treatment was 78 percent of the NSDUH estimate, and the count of past year admissions was 56 percent of the N-SSATS estimate only. The TEDS undercount varied by State and seemed to be most severe among inpatient hospital treatment populations and among individuals with low socioeconomic status. N-SSATS and NSDUH estimates of the single-day count were generally similar, but the NSDUH data showed a higher proportion of clients receiving treatment for alcohol use. Because of the considerable variation between the three data systems in terms of coverage, methods and timing of data collection, definitions, and information collected, each system has its own strengths and limitations that must be considered when deciding which one should be used in order to address specific policy or research questions.

## **Challenges of using prediction models to produce nationally representative estimates of serious mental illness**

**CITATION:** Hedden, S., Gfroerer, J., Bose, J., Kott, P., Liao, D., & Colpe, L. (2013, November). *Challenges of using prediction models to produce nationally representative estimates of serious mental illness*. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC.

**PURPOSE/OVERVIEW:** Using the National Survey on Drug Use and Health (NSDUH), the Substance Abuse and Mental Health Services Administration (SAMHSA) has been producing model-based prevalence estimates of past year serious mental illness (SMI) and any mental illness (AMI) among adults aged 18 or older since 2008. This paper provides a general overview of the methods used to produce estimates of SMI among adults using NSDUH and describes the revisions that were made to the estimation procedures to produce more accurate estimates.

**METHODS:** In addition to the main NSDUH interview, a nationally representative subsample of NSDUH adult respondents who completed the interview in English was asked to complete a follow-up interview on mental health for the 2008 to 2012 survey years (i.e., the Mental Health Surveillance Study [MHSS] clinical follow-up interviews). Using the clinical interview data collected in 2008, a statistical model and cut point were used to produce estimates of SMI among all adults. By the end of 2012, the combined sample from the MHSS collection was approximately 5,500 adults. Using the larger sample, two components of the estimation methods were investigated to determine whether revisions to the 2008 estimation methodology could be made that would increase the accuracy of the SMI estimates. Three criteria were used for model selection and to determine whether a revised model would be an improvement over the 2008 model: (1) minimization of misclassification, (2) minimization of subpopulation bias, and (3) model parsimony.

**RESULTS/CONCLUSIONS:** The model-based methodology allows SAMHSA to take advantage of the large NSDUH sample size for estimating SMI and AMI while using the small clinical interview sample to create a prediction model that links the short-scale data from the full NSDUH sample to the actual diagnostic assessments. Several issues were considered when developing these methods. Particularly discussed in this paper were the challenges of producing unbiased estimates and trends for the overall adult population and for subpopulations and the error incurred when producing model-based estimates.

## **On moving to the use of a hybrid sampling frame in the National Survey on Drug Use and Health: Motivations and challenges**

**CITATION:** Hughes, A., Bose, J., Shook-Sa, B., & Morton, K. (2013, November). *On moving to the use of a hybrid sampling frame in the National Survey on Drug Use and Health: Motivations and challenges*. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC.

**PURPOSE/OVERVIEW:** The sampling frame used in the National Survey on Drug Use and Health (NSDUH) is constructed by listers who visit each selected area segment and conduct a complete enumeration of dwelling units that are eligible for selection. The half-open interval (HOI) procedure is used by field interviewers during the data collection phase to identify and include dwelling units that were missed during the listing operation. Address-based sampling (ABS) has become a viable alternative to traditional approaches for developing sampling frames and can significantly reduce costs associated with the listing operation. This paper evaluates the coverage, cost, and implementation of a hybrid sampling frame as an alternative to using traditional field enumeration (FE) methods in NSDUH sample area segments where ABS would be employed in most area segments while the traditional listing method would continue to be used in areas where ABS coverage is low.

**METHODS:** The authors summarized the overall NSDUH sample design. Specifically, they reviewed the 2005-2013 NSDUH sample design, the new 2014 NSDUH sample design, the sampling framing procedures about how to identify missed addresses in ABS through examples of different Check for Housing Units Missed (CHUM) components, an expansion of the coverage of ABS via the No-Stat file that supplements the United States Postal Service

Computerized Delivery Sequence (USPS CDS) file with approximately 7 million predominantly rural addresses not found on the CDS file, the use of the hybrid frame, and the estimated coverage of NSDUH dwelling units of the hybrid frame.

RESULTS/CONCLUSIONS: SAMHSA decided not to pursue the use of a hybrid sampling frame in NSDUH for the following reasons: (1) potential increased field interviewer burden, (2) different procedures required to identify missed dwelling units in ABS and FE area segments, (3) limited ABS coverage in rural areas, and (4) limited coverage of the group quarters population. Additional field studies may be conducted to further test uniform procedures for identifying and adding missed dwelling units in both ABS and FE area segments.

## **2012 National Survey on Drug Use and Health: A revised strategy for estimating the prevalence of mental illness**

CITATION: Kott, P., Hedden, S., Aldworth, J., Bose, J., Chromy, J., Gfroerer, J., & Liao, D. (2013, October). *2012 National Survey on Drug Use and Health: A revised strategy for estimating the prevalence of mental illness* (RTI 0212800.002.120.008.002.004, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283201000003C). Research Triangle Park, NC: RTI International.

PURPOSE/OVERVIEW: The Substance Abuse and Mental Health Services Administration (SAMHSA) conducts the National Survey on Drug Use and Health (NSDUH) and publishes annual estimates of the prevalence of serious mental illness (SMI) and any mental illness (AMI) among adults aged 18 or older at the national, State, and substate level and within demographic groups. The purpose of this technical report was to summarize current and prior approaches and to document the research conducted to revise the 2012 strategy for estimating the prevalence of mental illness.

METHODS: Starting in 2008, SAMHSA added the Mental Health Surveillance Study (MHSS) to NSDUH. Until 2012, this included a clinical follow-up study of a sample of adult NSDUH respondents. From 2008 through 2011, SMI and AMI estimates were based on a statistical cut point model developed from 2008 data. The model used individual responses within NSDUH to the Kessler-6 (K6) scale and the abbreviated World Health Organization Disability Assessment Schedule (WHODAS) scale as indicators of SMI. The same model with a lower cut point was used to generate adult prevalence estimates for AMI. The model was revised in 2012 to include additional predictors of past year suicidal thoughts, past year major depressive episode, and an adjusted age.

RESULTS/CONCLUSIONS: The revised 2012 model was adopted for the 2012 NSDUH estimates of SMI and AMI. For comparability in estimating recent trends, the 2008 to 2011 SMI and AMI estimates were updated using the 2012 model and estimation procedures.

## **CBHSQ Data Review: Comparison of NSDUH health and health care utilization estimates to other national data sources**

CITATION: Pemberton, M. R., Bose, J., Kilmer, G., Kroutil, L. A., Forman-Hoffman, V. L., & Gfroerer, J. C. (2013, September). *CBHSQ Data Review: Comparison of NSDUH health and health care utilization estimates to other national data sources*. Rockville, MD: Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality.

PURPOSE/OVERVIEW: In addition to collecting data on substance use and mental health in the United States, the National Survey on Drug Use and Health (NSDUH) also collects data on health conditions and health care utilization. It is important for users of these data to recognize how NSDUH estimates differ from prevalence estimates produced by other nationally representative data sources, which have various objectives and scope, sampling designs, and data collection procedures. This report compares specific health conditions, overall health, and health care utilization prevalence estimates from the 2006 NSDUH and other national data sources.

METHODS: Methodological differences among these data sources that may contribute to differences in estimates are described. In addition to NSDUH, three of the data sources use respondent self-reports to measure health characteristics and service utilization: the National Health Interview Survey (NHIS), the Behavioral Risk Factor Surveillance System (BRFSS), and the Medical Expenditure Panel Survey (MEPS). One survey, the National Health and Nutrition Examination Survey (NHANES), conducts initial interviews in respondents' homes, collecting further data at nearby locations. Five data sources provide health care utilization data extracted from hospital records; these sources include the National Hospital Discharge Survey (NHDS), the Nationwide Inpatient Sample (NIS), the Nationwide Emergency Department Sample (NEDS), the National Hospital Ambulatory Medical Care Survey (NHAMCS), and the Drug Abuse Warning Network (DAWN). Several methodological differences that could cause differences in estimates are discussed, including type and mode of data collection; weighting and representativeness of the sample; question placement, wording, and format; and use of proxy reporting for adolescents.

RESULTS/CONCLUSIONS: The lifetime estimate of diabetes among adults from NSDUH (7.7 percent) did not differ from estimates found in the NHIS, NHANES, BRFSS, and MEPS. The lifetime estimate of asthma among adults from NSDUH (10.7 percent) was similar to the estimate from the NHIS (11.0 percent); estimates from other sources ranged from 9.6 to 14.2 percent. NSDUH's lifetime estimates of stroke and high blood pressure among adults were both lower than the estimates from the NHIS, NHANES, and MEPS, and considerable variation was noted between surveys in the rate of lifetime heart disease. Estimates of past year inpatient hospitalization among adults did not differ significantly between NSDUH and NHANES, but NSDUH's estimate was significantly higher than the estimates derived from the NHIS and MEPS. For both adults and adolescents, NSDUH's estimates of receiving treatment in an emergency department in the past year were higher than the estimates from other surveys. Demographic differences in the prevalence of chronic health conditions and health care utilization were similar across multiple surveys. Given all of the methodological differences among these data sources, the similarities among some estimates were noteworthy.

## National Survey on Drug Use and Health: Weighting assessment report

CITATION: Sathe, N., Chen, P., Dai, L., Laufenberg, J., Gordek, H., & Cribb, D. (2013, August). *National Survey on Drug Use and Health: Weighting assessment report* (RTI 0211838.108.006.010 and 0212800.001.108.006.010, prepared for the Substance Abuse and Mental Health Services Administration under Contract Nos. HHSS283200800004C and HHSS283201000003C). Research Triangle Park, NC: RTI International.

**PURPOSE/OVERVIEW:** This report summarizes the weighting assessment methods study for the National Survey on Drug Use and Health (NSDUH). The study's primary purpose was to investigate ways to streamline the current weighting process used to calibrate NSDUH's analysis weights without significantly affecting the estimates (totals and prevalence rates) produced by the current weighting procedure. Secondary goals of the study were to determine the effect on estimates of altering the age and race variables used during weight calibration and to investigate the utility of adding an educational attainment variable as a control variable in the final poststratification adjustment step. Also included in the study was an interviewer level of effort (LOE) investigation to determine whether paradata could be used to help reduce nonresponse bias in the survey estimates.

**METHODS:** The study involved 10 tasks: (1) removal of the screener dwelling unit-level poststratification (DUPS) adjustment, (2) removal of the selected person-level poststratification (SELPS) adjustment, (3) removal of the screener DUPS and person-level poststratification (PRPS) adjustments, (4) removal of all intermediate steps except the final PRPS adjustment, (5) use of a single year of age for those aged 12 to 25 in the person-level nonresponse (PRNR) and PRPS adjustments, (6) separating Native Hawaiians or Pacific Islanders from Asians in the PRNR and PRPS adjustments, (7) use of fewer control variables in weight calibration, (8) use of an education variable in the PRPS adjustment, (9) use of an interview callback model (level of effort [LOE]), and (10) analysis of the combined effect of implementing Tasks 1, 2, 5, 6, and 8.

**RESULTS/CONCLUSIONS:** The results and conclusions for each of the 10 tasks were as follows:

Task 1. Removal of DUPS adjustment: The effect of removing DUPS on the standard error of estimates was mixed (depended on the outcome). However it would shorten the person-level weighting process and reduce the weight variation by removing one adjustment step.

Task 2. Removal of SELPS adjustment: Skipping the SELPS adjustment had slightly more impact on the estimated percentages than on the estimated totals. However, it would shorten the person-level weighting process and reduce the weight variation by removing one adjustment step.

Task 3. Removal of the screener DUPS and PRPS adjustment: This was an extension of Task 1. However, it affected a large number of estimates.

Task 4. Removal of all intermediate steps except the final PRPS adjustment: Although the empirical results for this task could lead to the consideration of eliminating all of the

intermediate weighting steps and going straight to the final poststratification step (PRPS), it cannot be theoretically or conceptually justified.

Task 5. Use of a single year of age for those aged 12 to 25 in the PRNR and PRPS adjustments: The impact on the estimated totals was larger than on the estimated percentages. It had minimal impact on the associated standard errors. In fact, a reduction in standard errors associated with estimates for age-specific subgroups was observed.

Task 6. Separating Native Hawaiians or Pacific Islanders from Asians in the PRNR and PRPS adjustments: A smaller impact on estimated percentages than on estimated totals was observed (for most outcomes), and it had almost no impact on the associated standard errors. In fact, a reduction in standard errors associated with estimates for race/ethnicity subgroups was observed.

Task 7. Use of fewer control variables in weight calibration: All current variables were significant.

Task 8. Use of an education variable in the PRPS adjustment: The impact on the estimated totals was slightly larger than on the estimated percentages. It had minimal effect on the associated standard errors. In fact, a reduction in standard errors associated with estimates for education-specific subgroups was seen.

Task 9. Use of an interview callback model (LOE): There was no uniformly best approach for reducing the effects of nonresponse on survey estimates. None of the four callback models outperformed the current GEM models in all situations.

Task 10. Combined effect of implementing Tasks 1, 2, 5, 6, and 8: The weighting process that incorporated all of the recommended changes retained the impact of those changes: (1) Many significant results occurred in the age-specific subgroups, as seen in Task 5; (2) some significant results for Native Hawaiians or Other Pacific Islanders, as seen in Task 6, were retained in the final weighting process; (3) some significant results for education subgroups, as seen in Task 8, were retained in the final weighting process; (4) the large relative differences observed from Tasks 1 and 2 in the race/ethnicity subgroups also were seen in the final weighting process; (5) no unexpected significant results were observed in the final weighting process. The majority of significant results in the final weighting process could be linked to one or more individual tasks. For the significant results that could not be linked to any of the five tasks, the absolute relative differences (ARDs) were small to moderate.

## **An investigation of decennial census effects on estimates of substance use and mental illness from the National Survey on Drug Use and Health (NSDUH)**

CITATION: Sathe, N., Chen, P., Hughes, A., Bose, J., Dai, L., & Foster, M. (2013, August). *An investigation of decennial census effects on estimates of substance use and mental illness from the National Survey on Drug Use and Health (NSDUH)*. Paper presented at the 2013 Joint Statistical Meetings of the American Statistical Association, Montreal, Canada.

**PURPOSE/OVERVIEW:** The National Survey on Drug Use and Health (NSDUH), an annual survey of the U.S. civilian, noninstitutionalized population aged 12 or older, is a major source of substance use and mental illness data. The 2010 estimates were produced using weights poststratified to 2010 population control totals (intercensal estimates) derived from the 2000 decennial census; however, the 2011 estimates were produced using weights poststratified to 2011 population control totals derived from the 2010 decennial census. This study was done to determine whether the change in the source of the control totals had an effect on the level of change observed between the 2010 and the 2011 estimates.

**METHODS:** To examine this "census effect," 2010 estimates were also produced using weights poststratified to 2010 population control totals derived from the 2010 decennial census, resulting in two sets of weights for use on the 2010 data. NSDUH estimates were compared using both sets of 2010 estimates along with the 2011 estimates.

**RESULTS/CONCLUSIONS:** Substance use estimates were more affected by the census effect than were the estimates of mental illness, and they were more notable for estimated totals compared with rates.

### **Extending the coverage of address-based sampling frames: Beyond the USPS computerized delivery sequence file**

**CITATION:** Shook-Sa, B. E., Currivan, D. B., McMichael, J. P., & Iannacchione, V. G. (2013). Extending the coverage of address-based sampling frames: Beyond the USPS computerized delivery sequence file. *Public Opinion Quarterly*, 77(4), 994-1005. doi:10.1093/poq/nft041

**PURPOSE/OVERVIEW:** Because of rising data collection costs associated with in-person surveys, address-based sampling (ABS) is being explored as an alternative to traditional field enumeration (FE). This has led surveys such as the National Survey on Drug Use and Health (NSDUH), which relies solely on FE, to investigate the use of ABS as a potential sampling frame. Before 2009, ABS frames were restricted to the Computerized Delivery Sequence (CDS) file, which the United States Postal Service (USPS) makes available through licensing agreements with qualified vendors. Research based on the CDS has found the coverage of ABS frames for in-person surveys to be sufficient in urban areas but problematic in rural areas. Since 2009, the USPS has also made available the No-Stat file. A supplement to the CDS file, the No-Stat file contains approximately 7 million predominantly rural addresses, most of which are not currently receiving mail directly and are not found in the CDS. One component of evaluating ABS for NSDUH was to explore the potential added coverage and the cost-effectiveness of using the No-Stat file as a supplemental address source.

**METHODS:** The authors selected a subsample of occupied housing units (households) eligible for NSDUH from quarter 2 of the 2010 NSDUH sample covering April through June 2010. The analytic goal was to estimate the household (HH) coverage provided by the CDS and No-Stat files by matching the addresses of each subsampled HH to the April 2010 files.

**RESULTS/CONCLUSIONS:** The No-Stat file provided an overall coverage rate of 1.2 percent compared with the 93.2 percent coverage provided by the CDS file. Although the overall

coverage provided by the No-Stat file was low, the No-Stat file did increase coverage in rural and rural/urban mix segments. It added 3.8 percent coverage to the ABS frame in rural segments, bringing rural coverage from 72.8 percent without the No-Stat file to 76.6 percent with the No-Stat file.

## **An empirical study to evaluate the performance of synthetic estimates of substance use in the National Survey on Drug Use and Health**

CITATION: Vaish, A., Sathe, N., Spagnola, K., Folsom, R., & Hughes, A. (2013, August). *An empirical study to evaluate the performance of synthetic estimates of substance use in the National Survey on Drug Use and Health*. Paper presented at the 2013 Joint Statistical Meetings of the American Statistical Association, Montreal, Canada.

PURPOSE/OVERVIEW: Small area estimation methods are used to produce State and substate estimates of substance use and mental disorders using data from the National Survey on Drug Use and Health (NSDUH). Design-based estimates could be used as an alternative because they are less expensive than small area estimates and take less time to produce. Thus, it is important to determine how the small area estimates compare with their design-based counterparts in terms of accuracy and precision. A previous study demonstrated that small area estimates were generally more precise than design-based estimates while exhibiting only small levels of bias. In this paper, those results are extended by conducting an additional simulation study to evaluate the performance of synthetic estimates. These estimates are commonly produced for small areas where no sample data can be obtained, and this study aimed to provide some guidance about the quality of such estimates.

METHODS: For this study, pooled 2009 and 2010 NSDUH data were used. In NSDUH, each of the eight large States (i.e., California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas) has 48 State sampling regions (SSRs), and each of the 48 SSRs is expected to have eight segments in a single year of NSDUH data. Out of eight segments, about half of the segments are used again in the next year for sampling, and the other half of the segments are replaced with new segments. Each of the eight large States was partitioned into two parts with  $n = \sim 3,600$  resulting in  $8 \times 2 = 16$  benchmark areas, and their design-based estimates were treated as "true value." After creating 16 benchmark areas, the authors created 16 pseudo substate areas with  $n = 225$  (approximately) in each of the 16 benchmark areas, resulting in  $16 \times 16 = 256$  pseudo substate areas. For producing small area estimates, the age group-specific (12 to 17, 18 to 25, 26 or older) respondent-level weights in each of the 256 pseudo substate areas were poststratified to the corresponding 1/16 of the benchmark area-level analysis weight totals. Small area estimates were produced for four outcome measures—past month binge alcohol use (BNGALC), past month cigarette use (CIGMON), past year cocaine use (COCYR), and past month marijuana use (MRJMON)—using a simpler version of NSDUH's current small area estimation models (i.e., only age group-specific pseudo substate area-level random effects were included in the logistic mixed models along with the current fixed-effects predictors). Using the fitted model, small area estimates and synthetic estimates for the 256 pseudo substate areas for the four outcome measures were produced. The corresponding design-based estimates (weighted averages) were also produced.

**RESULTS/CONCLUSIONS:** The synthetic estimates performed better than expected, especially for the higher prevalence outcome measures (BNGALC, CIGMON). Perhaps after adjusting for age, race, gender, and a few other covariates, not enough variability remained among the pseudo substate areas to benefit from the fitted substate area-level random effects. This will need to be investigated further. The confidence intervals for synthetic estimates were much wider than expected; that is, the simulation mean squared errors (MSEs) were much smaller than the model-based MSEs for the synthetic estimates. This will also need to be investigated further.

## **Assessing the relationship between interviewer effects and NSDUH data quality**

**CITATION:** Wang, K., Kott, P., & Moore, A. (2013, October 24). *Assessing the relationship between interviewer effects and NSDUH data quality* (RTI 0211838.108.006.020 and 0212800.001.108.006.020, prepared for the Substance Abuse and Mental Health Services Administration under Contract Nos. HHSS283200800004C and HHSS283201000003C). Research Triangle Park, NC: RTI International.

**PURPOSE/OVERVIEW:** For interviewer-administered surveys, interviewers can have a major effect on various measures of survey data quality, including unit response rates, item missing rates, estimates of uncertainty (variances), and the accuracy of survey estimates. Interviewers play a crucial role in gaining initial cooperation from respondents, then sustaining that cooperation throughout the interview. Interviewers are also responsible for administering the survey instrument, perhaps asking follow-up questions or probes when the respondent does not answer a question adequately. Finally, interviewers are responsible for recording responses into the data collection instrument. Decisions regarding the National Survey on Drug Use and Health (NSDUH) sample and survey design could affect the number of NSDUH field interviewers (FIs) and the composition of NSDUH FIs in terms of experience, the number of hours they work, the distances they travel, and so on. These factors have been shown to be related to key NSDUH outcome measures. To inform future decisions about NSDUH's survey and sample design, it is crucial to understand the relationships between interviewer characteristics and outcomes of interest, such as response rates and prevalence rates. Such knowledge then can be used not only to inform the degree to which changes in NSDUH's study design that affect interviewer composition and work practices may affect response rates, prevalence rates, and other outcome measures of interest, but also to direct efforts at identifying and removing the causes of different prevalence rates.

**METHODS:** The authors summarized existing research on the relationships between interviewer characteristics and data quality on NSDUH and described the variables used in the previous analyses. They reported on three repeated cross-sectional analyses of the effects of interviewer characteristics on measures of survey participation: screener contact, screener cooperation, and interview cooperation. They also conducted analyses on the relationships between interviewer characteristics and substantive items of interest from the survey, namely, measures of substance use and mental health outcomes. Additionally, they described the results of two cohort analyses that were carried out to assess whether the relationship between survey outcomes and interviewer experience uncovered from the previous analyses was due to the gaining of experience or the

nature of interviewers who eventually gained experience (even after using regression techniques to control for their caseloads and other characteristics).

**RESULTS/CONCLUSIONS:** Interviewer experience remained positively correlated with response rates and negatively associated with self-reported substance use. However, evidence suggested that the association between experience and substance use reporting had declined over time. Furthermore, the pattern of declining effects was more conspicuous with measures based on the cumulative interview count than with quarters or years of experience. Analyses in which interviewer experience was measured by length of time on the job, such as quarters or years worked on NSDUH since 1999, yielded different results. Measures based on the cumulative interview count produced statistically significant effects on substance use outcomes in the earlier years of the study, whereas the duration-based experience measures (quarters worked, years worked) were not statistically significant predictors for as many outcomes. For the 10 substance use measures examined in this study, the effects of a continuous measure of the cumulative interview count were statistically significant in 2002, and the largest effects (in terms of magnitude) were observed in that year than in any other year. In contrast, the effects of the quarter-based measure of experience were statistically significant for only four measures, and the effects of the year-based measure were statistically significant for only three measures in 2002.

# 2014

## **CBHSQ Data Review: Comparing and evaluating substance use treatment utilization estimates from the National Survey on Drug Use and Health and other data sources**

CITATION: Batts, K., Pemberton, M., Bose, J., Weimer, B., Henderson, L., Penne, M., Gfroerer, J., Trunzo, D., & Strashny, A. (2014, April). *CBHSQ Data Review: Comparing and evaluating substance use treatment utilization estimates from the National Survey on Drug Use and Health and other data sources*. Rockville, MD: Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality.

**PURPOSE/OVERVIEW:** This report presents an evaluation of the coverage, overlap, biases, strengths, and weaknesses of three data sources about the receipt of specialty substance use treatment: the National Survey on Drug Use and Health (NSDUH), the National Survey of Substance Abuse Treatment Services (N-SSATS), and the Treatment Episode Data Set (TEDS). NSDUH is an annual survey of a representative sample of individuals aged 12 or older at their places of residence that includes questions on the receipt of substance use treatment. N-SSATS is an annual census of all known drug and alcohol use treatment facilities in the United States. TEDS is a compilation of data about the demographic and substance use characteristics of admissions to and discharges from substance use treatment.

**METHODS:** Methodological differences among these data sources that could contribute to the differences in estimates were described. Specialty substance use treatment measures included in this paper's comparisons were the numbers and characteristics of individuals treated in a given year, single-day treatment counts, numbers of admissions in a given year, and estimates of the numbers of individuals who needed substance use treatment but did not receive it.

**RESULTS/CONCLUSIONS:** NSDUH estimates of individuals treated in a given year were significantly higher than the estimates from TEDS. Single-day treatment counts from NSDUH were similar to those from N-SSATS, and both were significantly higher than those from TEDS. N-SSATS counts of annual admissions were significantly higher than counts derived from TEDS. The consistently lower counts in TEDS seemed to be due to coverage differences in the three data systems. TEDS is mostly limited to those individuals whose treatment was publicly funded, whereas N-SSATS includes a census of all facilities regardless of funding, and NSDUH includes individuals who are treated in both privately and publicly funded facilities. Precise agreement among the data sources was not generally expected, and this lack of agreement did not reduce the importance of these studies in their contribution to the understanding of specialty substance use treatment in the United States. The analyses presented in this paper provide a basis for improving the interpretation of results from these studies and facilitate developing clear guidance for future analyses to better answer some basic questions about substance use treatment, such as how many individuals receive treatment in a year, how large is the gap between treatment received and treatment needed, and how have the numbers of individuals receiving and needing treatment changed over time.

## **Discrepancies in prevalence estimates in two national surveys for nonmedical use of a specific opioid product versus any prescription pain reliever**

CITATION: Biondo, G., & Chilcoat, H. D. (2014). Discrepancies in prevalence estimates in two national surveys for nonmedical use of a specific opioid product versus any prescription pain reliever. *Drug and Alcohol Dependence*, 134, 396-400.

BACKGROUND: The need to understand trends in the nonmedical use of prescription pain relievers as a class, as well as specific opioid products, is growing. Surveys such as Monitoring the Future (MTF) study and the National Survey on Drug Use and Health (NSDUH) are important tools for understanding trends in the nonmedical use of prescription drugs and the use of illegal drugs. This report compares discrepancies in prevalence estimates between these surveys for a specific opioid product (oxycodone) relative to other drugs.

METHODS: Trends in past year use of marijuana, cocaine, and nonmedical use of oxycodone and any prescription pain reliever were estimated for each survey from 2005 to 2010 for adolescents in the 12th grade. The proportion of nonmedical pain reliever users who misused oxycodone was estimated for each survey. In NSDUH, the term OxyContin<sup>®</sup> is used and refers to controlled-release oxycodone hydrochloride.

RESULTS/CONCLUSIONS: Prevalence of the nonmedical use of oxycodone in the past year was steady over time for both surveys, but it was 2.5 to 3 times higher in the MTF compared with NSDUH. Trends in the prevalence of marijuana and cocaine use were similar across both surveys, although prevalence estimates for each were on average 18 percent higher in the MTF. In contrast, prevalence estimates for any nonmedical prescription pain reliever use were on average 15 percent lower in the MTF. The proportion of nonmedical prescription pain reliever users who used oxycodone was 42 percent in the MTF versus 19 percent in NSDUH. The discrepancy between surveys in prevalence estimates for the nonmedical use of oxycodone exceeded those for other drugs, pointing to the importance of visual aids and items used to measure the nonmedical use of specific products.

## **Monitoring of non-cigarette tobacco use using Google Trends**

CITATION: Cavazos-Rehg, P. A., Krauss, M. J., Spitznagel, E. L., Lowery, A., Gruzca, R. A., Chaloupka, F. J., & Bierut, L. J. (2014). Monitoring of non-cigarette tobacco use using Google Trends. *Tobacco Control*, doi:10.1136/tobaccocontrol-2013-051276

PURPOSE/OVERVIEW: Google Trends is an innovative monitoring system with unique potential to monitor and predict important phenomena that may be occurring at a population level. The authors sought to validate whether Google Trends can additionally detect regional trends in youth and adult tobacco use.

METHODS: The authors compared 2011 Google Trends relative search volume data for cigars, cigarillos, little cigars, and smokeless tobacco with State prevalence of youth (grades 9 to 12) and adult (age 18 or older) use of these products using data from the 2011 Youth Risk Behavior Survey, which is a component of the Youth Risk Behavior Surveillance System (YRBSS), and

the 2010-2011 National Survey on Drug Use and Health (NSDUH), respectively. They used the Pearson correlation coefficient to measure the associations.

**RESULTS/CONCLUSIONS:** The authors found significant positive correlations between State Google Trends cigar relative search volume and the prevalence of cigar use among youths ( $r = 0.39$ ,  $R^2 = 0.154$ ,  $p = 0.018$ ) and adults ( $r = 0.49$ ,  $R^2 = 0.243$ ,  $p < 0.001$ ). Similarly, the authors found that the correlations between State Google Trends smokeless tobacco relative search volume and the prevalence of smokeless tobacco use among youths and adults were both positive and significant ( $r = 0.46$ ,  $R^2 = 0.209$ ,  $p = 0.003$  and  $r = 0.48$ ,  $R^2 = 0.226$ ,  $p < 0.001$ , respectively). The results of this study validated that Google Trends has the potential to be a valuable monitoring tool for tobacco use. The near real-time monitoring features of Google Trends may complement traditional surveillance methods and lead to faster and more convenient monitoring of emerging trends in tobacco use.

## **National Survey on Drug Use and Health: 2012 Questionnaire Field Test final report**

**CITATION:** Center for Behavioral Health Statistics and Quality. (2014). *National Survey on Drug Use and Health: 2012 Questionnaire Field Test final report*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

**PURPOSE/OVERVIEW:** The Substance Abuse and Mental Health Services Administration's (SAMHSA's) Center for Behavioral Health Statistics and Quality (CBHSQ) is planning to implement changes related to a partial redesign of the National Survey on Drug Use and Health (NSDUH). These changes include use of a new sample design in 2014 and a limited update to the interview questionnaire in 2015. The new sample design will allow for continued national-, State-, and substate-level estimation comparable with estimation from previous surveys. The sample design's improved efficiency will result in significant cost savings. The primary change to the questionnaire is an updated set of prescription drug modules, which will include current prescription drugs and incorporate a new questionnaire structure. Other planned changes to the questionnaire include a revised health module that contains new questions about drug and alcohol screening by primary care physicians. These changes will seek to achieve three main goals: (1) to revise the questionnaire to address changing policy and research data needs, (2) to modify the survey methodology to improve the quality of estimates and the efficiency of data collection and processing, and (3) to maintain trends in core substance use estimates across survey years. The 2012 Questionnaire Field Test (QFT) was meant to test the revisions to the questionnaire and protocols, and this report describes the data collection and analytic methods and results of the 2012 QFT, including comparisons of selected QFT estimates with current and comparable NSDUH data and other data sources.

The primary goal of the field test was to measure the total effect on NSDUH estimates from all changes to the protocol planned for the 2015 redesign, using multiple indicators. This report summarized how the QFT was conducted and the results obtained to address the four main research questions as follows:

1. To what extent do the planned changes in the protocol influence data quality as measured by unit nonresponse, item nonresponse, imputation rates, and other indicators of data quality?
2. To what extent does the redesigned protocol influence the overall timing of the full interview, the section timing for revised modules, and the screener timing, including the new field observation questions?
3. What measurable implications, if any, for the general feasibility of the redesigned protocol were obtained from field observations, field interviewer (FI) debriefing items, equipment surveys, or focus groups with QFT interviewers?
4. To what extent are the planned changes in the protocol associated with any increases or decreases in the reporting of core substance use, methamphetamine, prescription drugs, or noncore items?

**METHODS:** Similar to the NSDUH main study, the respondent universe for the QFT was the civilian, noninstitutionalized population aged 12 or older. In order to control costs, individuals residing in Alaska and Hawaii, as well as those who were not able to complete the interview in English, were excluded from the QFT. A total of 5,358 dwelling units were sampled and yielded 2,044 completed interviews. Estimates for survey items from the QFT were compared with data from the 2011 NSDUH and from quarters 3 and 4 from the 2012 NSDUH (which also excluded respondents from Alaska and Hawaii and those who were not able to complete the interview in English). Estimates for some items were also compared with external benchmark data sources, such as the National Health Interview Survey (NHIS), the American Community Survey (ACS), and the Current Population Survey (CPS). In addition, data from the National Health and Nutrition Examination Survey (NHANES), the National Ambulatory Medical Care Survey (NAMCS) and the hospital outpatient clinic component of the National Hospital Ambulatory Medical Care Survey (NHAMCS) were used to assess the ranking of mentions of prescription drugs in the QFT. Data on interview timing and missing data rates from the QFT were compared with data from the 2011 NSDUH and quarters 3 and 4 from the 2012 NSDUH. Finally, qualitative information was gathered from interviewer debriefing items, field observations, and focus groups with interviewers.

**RESULTS/CONCLUSIONS:** Response rates in the QFT were lower than those in the 2011 NSDUH as well as quarters 3 and 4 from the 2012 NSDUH. This may have been due to the short field period of the QFT relative to the main survey and limited flexibility to assign interviewers to particular cases in the QFT.

Some items that were moved from computer-assisted personal interviewing (CAPI) to audio computer-assisted self-interviewing (ACASI) administration in the QFT questionnaire produced significantly higher missingness rates than the 2011 and 2012 quarters 3 and 4 comparison data, particularly for items on health insurance and income.

The overall mean interview time for the QFT interviews was lower than the mean times for the 2011 and 2012 quarters 3 and 4 comparison interviews, but additions and revisions to the hallucinogens, inhalants, and prescription drug sections in the partially redesigned QFT questionnaire contributed to higher administration times for the core substance use modules

compared with the 2011 and 2012 quarters 3 and 4 comparison interviews. The redesigned prescription drug modules had greater QFT administration times for these modules, primarily in the pain relievers module. Lower mean times for several back-end demographic sections (including employment, income, and administrative residual times) for the QFT interviews contributed significantly to the lower overall interview times compared with the 2011 and 2012 quarters 3 and 4 comparison interviews.

Most estimates for prevalence rates for core substances (cigarettes, alcohol, marijuana, cocaine, and heroin) in the QFT appeared similar to estimates from the 2011 and 2012 quarters 3 and 4 comparison samples.

The lifetime estimate for methamphetamine use among individuals aged 12 or older was higher in the QFT than in the comparison data from 2011 and 2012 quarters 3 and 4. Prescription drug estimates for lifetime misuse among all individuals aged 12 or older were lower in the QFT data than in the 2011 and 2012 quarters 3 and 4 comparison data for pain relievers and tranquilizers. Estimates of past year misuse for pain relievers, OxyContin<sup>®</sup>, and sedatives among individuals aged 12 or older were higher for the QFT than for the 2011 and 2012 quarters 3 and 4 comparison data.

Estimates for some items in the QFT that were administered in ACASI that were administered in CAPI in the main study were higher in the QFT than in the 2011 and 2012 comparison data. These included participation in government assistance programs, receiving supplemental security income, and participating in food stamp programs.

## **National Survey on Drug Use and Health: 2013 Dress Rehearsal final report**

CITATION: Center for Behavioral Health Statistics and Quality. (2014). *National Survey on Drug Use and Health: 2013 Dress Rehearsal final report*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

PURPOSE/OVERVIEW: The Substance Abuse and Mental Health Services Administration's (SAMHSA's) Center for Behavioral Health Statistics and Quality (CBHSQ) is planning to implement changes related to a partial redesign of the National Survey on Drug Use and Health (NSDUH). These changes include use of a new sample design in 2014 and a limited update to the interview questionnaire in 2015. The new sample design will allow for continued national-, State-, and substate-level estimation comparable with estimation from previous surveys. The sample design's improved efficiency will result in significant cost savings. The primary change to the questionnaire is an updated set of prescription drug modules, which will include current prescription drugs and incorporate a new questionnaire structure. Other planned changes to the questionnaire include a revised health module that contains new questions about drug and alcohol screening by primary care physicians. These changes will seek to achieve three main goals: (1) to revise the questionnaire to address changing policy and research data needs, (2) to modify the survey methodology to improve the quality of estimates and the efficiency of data collection and processing, and (3) to maintain trends in core substance use estimates across survey years.

This report summarizes the data collection and analytic methods and results for NSDUH's 2013 Dress Rehearsal (DR). Using multiple indicators and data sources, the primary goal of the DR was to measure the total effect on NSDUH estimates and outcomes from all changes to the materials, questionnaire, and procedures planned for the 2015 partial redesign. This report summarizes how the DR was conducted and the results obtained to address the five main research questions as follows:

1. What do assessments of the DR protocol obtained from equipment surveys, debriefing questions, debriefing calls, and field observations of field interviewers (FIs) indicate about the likely effectiveness of the 2015 partial redesign protocol?
2. What impact does the redesigned protocol, including revisions made to the DR questions or protocol based on Questionnaire Field Test (QFT) experiences or results, have on the overall interview timing and module timings across age groups?
3. Does the DR protocol, including changes made from the QFT protocol, meet similar data quality standards as the QFT data collection and the current NSDUH main study, as measured by unit nonresponse, item missingness rates, imputation rates, and other indicators of data quality?
4. Does the DR protocol produce any significant differences in key estimates with the QFT and the main study comparison data, both for all respondents and across age groups and for both English-language and Spanish-language interviews?
5. Does the DR protocol produce any significant differences in key estimates relative to estimates from other surveys or other sources of data?

**METHODS:** Similar to the NSDUH main study, the respondent universe for the DR was the civilian, noninstitutionalized population aged 12 or older. In order to control costs, individuals residing in Alaska and Hawaii were excluded from the DR. A total of 4,392 dwelling units were sampled and yielded 2,087 completed interviews. The respondent sample was allocated to three age groups in the following proportions: 25 percent aged 12 to 17, 25 percent aged 18 to 25, and 50 percent aged 26 or older. This allocation matched the allocation for the QFT and 2014 NSDUH but differed from the allocation for the 2012 and 2013 NSDUHs.

One major difference between the QFT and the DR was the inclusion of Spanish-language interviews in the DR sample in order to allow for an assessment of the redesigned Spanish-language instrument. To achieve a higher yield of Spanish-language interviews in the DR sample than what would be observed with a probability proportional to size (PPS) sample, a special certainty stratum was created for State sampling regions (SSRs) with a historically high percentage of interviews conducted in Spanish. SSRs that had 10 percent or more of their 2011 NSDUH interviews conducted in Spanish were assigned to the certainty stratum.

Estimates for survey items from the combined DR and QFT samples (for non-Hispanic respondents who completed the interview in English) were compared with comparable data from the 2012 NSDUH and from quarters 3 and 4 from the 2013 NSDUH. Outcomes and estimates from the Spanish-language interviews in the DR were compared with Spanish-language interviews from the 2012 main study and 2013 quarters 3 and 4 main study.

Estimates for some items were also compared with external benchmark data sources, including the National Health Interview Survey (NHIS), the National Health and Nutrition Examination Survey (NHANES), the American Community Survey (ACS), the Current Population Survey (CPS), the General Social Survey (GSS) and the National Survey of Family Growth (NSFG). Data on interview timing and missing data rates from the QFT were compared with data from the 2011 NSDUH and quarters 3 and 4 from the 2012 NSDUH. Finally, qualitative information was gathered from interviewer debriefing items, field observations, and focus groups with interviewers.

**RESULTS/CONCLUSIONS:** Screening and interview response rates in the DR were lower than those in the 2012 NSDUH comparison sample and similar to those obtained in the QFT.

Overall, item missingness rates and variable imputation rates examined in the DR results were similar to the QFT results. Items that were moved from computer-assisted personal interviewing (CAPI) to audio computer-assisted self-interviewing (ACASI) administration in the QFT and DR questionnaires produced significantly higher missingness rates than in the main study comparison data.

Overall interview times were lower or similar for DR interviews compared with interviews from the 2012 and 2013 quarters 3 and 4 comparison data, and the QFT interviews for most age groups. Spanish-language DR respondents aged 12 or older, however, had higher overall interview times when compared with the Spanish-language 2012 respondents and the Spanish-language 2013 quarters 3 and 4 respondents. Interview times for Spanish-language DR data were much higher than either the Spanish-language 2012 main study or the Spanish-language 2013 quarters 3 and 4 interviews for respondents aged 65 or older; however, despite this larger difference in average times for respondents aged 65 or older, the overall timing differences for Spanish-language interviews were not of a large magnitude.

Similar to what was found in the QFT, estimates of *lifetime* misuse of prescription drugs were lower in the DR sample than in the NSDUH comparison data, but estimates of past year misuse were higher in both the QFT and combined QFT-DR data relative to the corresponding NSDUH comparison datasets.

For items on receipt of income from government assistance programs, participation in food stamp programs, employment and health insurance, some of the key differences in estimates observed between the QFT data, main study comparison data, and external data sources were observed for the DR data. The majority of these observed differences suggested that the DR sample was comprised of a higher proportion of respondents with lower socioeconomic status.

## **National Survey on Drug Use and Health: Investigation into text to speech software**

CITATION: Center for Behavioral Health Statistics and Quality. (2014). *National Survey on Drug Use and Health: Investigation into text to speech software*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

**PURPOSE/OVERVIEW:** Implementing text to speech (TTS) technology in the audio computer-assisted self-interviewing (ACASI) module of the National Survey on Drug Use and Health (NSDUH) interview offers an opportunity for work process efficiencies and cost savings in NSDUH's ACASI development. Although TTS software cannot match the audio quality of human voice recording, the evaluation presented in this report, along with the experience of both the National Survey of Family Growth (NSFG) and the Population Assessment of Tobacco and Health (PATH) survey, indicate that it was of sufficient quality to serve as a replacement for human voice recordings in ACASI. This report summarizes a follow-up investigation of TTS software, assesses commercially available TTS products that are potentially viable for use on NSDUH, and presents the pros and cons of utilizing each on NSDUH.

**METHODS:** The investigation was conducted in three phases: Phase 1 focused on researching different TTS systems and identifying products suitable for further evaluation; Phase 2 consisted of an evaluation of selected products; and Phase 3 focused on the development of the report and a recommendation regarding which of the evaluated products was most promising for NSDUH's ACASI modules.

**RESULTS/CONCLUSIONS:** The evaluation results indicated that a dynamic implementation offered a higher quality audio experience than the static implementation largely because of the elimination of audio files. Therefore, if TTS software is adopted for use with NSDUH, a dynamic implementation mode should be pursued over a static mode. With respect to the TTS products evaluated, Microsoft's Speech Platform was recommended. Considering the evaluated TTS products in both English and Spanish, the Microsoft product was ranked the highest by the evaluation team. Also, the Microsoft product offered a significant advantage over NeoSpeech because it was freely available and required no licensing agreement or user fees.

## **Primary measures of dependence among menthol compared to non-menthol cigarette smokers in the United States**

**CITATION:** Curtin, G. M., Sulsky, S. I., Van Landingham, C., Marano, K. M., Graves, M. J., Ogden, M. W., & Swauger, J. E. (2014, May 20). Primary measures of dependence among menthol compared to non-menthol cigarette smokers in the United States. *Regulatory Toxicology and Pharmacology: RTP*. doi:10.1016/j.yrtph.2014.05.011

**PURPOSE/OVERVIEW:** Previously published studies provide inconsistent evidence on whether menthol in cigarettes is associated with increased dependence. A few national studies, including the National Health and Nutrition Examination Survey (NHANES), National Survey on Drug Use and Health (NSDUH), National Health Interview Survey (NHIS), and Tobacco Use Supplement to the Current Population Survey (TUS-CPS), collect data on current cigarette type preference and primary measures of dependence. These studies allow researchers to examine whether menthol smokers are more dependent than nonmenthol smokers.

**METHODS:** The authors used the data from several surveys (NHANES, NSDUH, and NHIS and TUS-CPS) to get the number of cigarettes smoked per day among menthol and nonmenthol cigarette smokers and the time to first cigarette after waking and Heaviness of Smoking Index. To properly estimate variances, they weighted the dataset to the U.S. population and used survey

statistics to account for the complex sample designs and conducted the analyses with the goal of maximizing comparability across datasets.

**RESULTS/CONCLUSIONS:** Analyses based on combined data from multiple administrations of each of these four nationally representative surveys, using three definitions for current smokers (i.e., smoked 1 day, 10 days, and daily during the past month), consistently demonstrated that menthol smokers did not report smoking more cigarettes per day than nonmenthol smokers. Moreover, two of the three surveys that provided data on the time to first cigarette after waking indicated no difference in urgency to smoke among menthol compared with nonmenthol smokers, while the third suggested menthol smokers may experience a greater urgency to smoke; estimates from all three surveys indicated that menthol versus nonmenthol smokers did not report a higher Heaviness of Smoking Index. Collectively, these findings indicated no difference in dependence among U.S. smokers who use menthol compared with nonmenthol cigarettes.

### **CBHSQ Data Review: Arrestee substance use: Comparison of estimates from the National Survey on Drug Use and Health and the Arrestee Drug Abuse Monitoring Program**

**CITATION:** Lattimore, P. K., Steffey, D. M., Gfroerer, J., Bose, J., Pemberton, M. R., & Penne, M. A. (2014, August). *CBHSQ Data Review: Arrestee substance use: Comparison of estimates from the National Survey on Drug Use and Health and the Arrestee Drug Abuse Monitoring Program*. Rockville, MD: Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality.

**PURPOSE/OVERVIEW:** To address the treatment needs of arrestees and to develop proper programs and policies for dealing with drug use among arrestees, it is important for policymakers to have accurate information on the substance use treatment needs among the arrestee population. Two primary sources of such data are available: the Arrestee Drug Abuse Monitoring (ADAM) Program and the National Survey on Drug Use and Health (NSDUH).

ADAM was established by the National Institute of Justice and was fully implemented in 2000 in 35 communities around the United States. It consisted of interviews with and drug tests of arrestees in local jails within 48 hours of their being booked and was expanded to 39 communities including 41 counties in 2003. The survey captured information from respondents about substance use, including self-reported use and urine test results, receipt of substance use treatment, and drug market participation, including information about how illicit drugs were acquired and whether drugs were obtained in cash transactions or through other exchanges. ADAM data collection ended after 2003; however, a similar program, ADAM II, samples arrestees and conducts interviews and urinalysis drug screens in booking facilities in 10 communities nationwide. Data from ADAM and ADAM II sites were collected during nonrandomly selected 1- to 4-week data collection periods and are not generalizable to the sites or to the Nation as a whole. Thus, the calculation of site-specific or national estimates for substance use among arrestees is not straightforward for ADAM data. In addition to questions about alcohol use and illicit drug use, NSDUH also asks about arrests in the past 12 months and whether a respondent is on probation or parole. This CBHSQ Data Review attempts to evaluate the coverage and quality of drug use estimates of arrestees in the NSDUH data based on

comparisons with ADAM data and Uniform Crime Reporting (UCR) data. The Federal Bureau of Investigation's UCR data provide annual estimates of arrests at the county level.

**METHODS:** The analyses compared results of substance use in self-reports by adult males from the 2003 ADAM data collected within 48 hours of arrest and by adult males who reported at least one arrest in the previous year from the NSDUH data collected in the 39 ADAM sites. NSDUH data from 2002 to 2008 were combined for this analysis in order to yield a large enough sample for reliable estimates, resulting in a total of approximately 1,800 NSDUH adult male past year arrestees in these 39 sites. For this study, ADAM subjects were classified as NSDUH eligible or NSDUH ineligible based on their response to an ADAM interview question on where they had lived most of the time during the 30 days prior to arrest. Those who reported living in a house, mobile home, apartment, residential hotel, rooming house, dormitory, group home, student housing, or a shelter were considered to be eligible for inclusion in NSDUH and therefore were treated as NSDUH eligible for the purposes of the comparative analyses.

Of the 22,903 adult male ADAM cases in 2003, 20,457 were classified as NSDUH eligible, and 1,953 were classified as NSDUH ineligible. ADAM respondents with unknown NSDUH eligibility ( $n = 493$ ) were excluded from the NSDUH-eligible and the NSDUH-ineligible estimates.

**RESULTS/CONCLUSIONS:** NSDUH data provide generalizable national-level estimates for arrestees and a depth of contextual information about their drug use, mental health conditions, health status, and demographics. Despite difficulties in making comparisons with ADAM data because of ADAM's lack of generalizability, it appears that in addition to known undercoverage due to the NSDUH eligibility rules (about 9 percent of arrests), there was some additional undercoverage, which was estimated to be about 15 percent. Also, substance use and dependence and abuse rates from NSDUH appeared to be lower for reasons beyond differences in coverage, such as the difference in the reference period (i.e., past month in NSDUH vs. the 30 days prior to the arrest in ADAM).

# Author and Subject Index

## Author Index

Abbacchi, A. M.....	148
Abelson, H. I.....	7
Aldworth, J.....	113, 127, 134, 141, 151, 163, 166, 170, 179, 201
Anthony, J. C.....	37, 71, 105
Aquilino, W. S.....	64
Barchha, N.....	155
Barker, P.....	57, 67, 73, 80, 83, 104, 107, 118, 134, 139, 161, 163, 166, 169, 173, 179, 188
Barnett-Walker, K.....	163
Barnwell, B. G.....	74
Batts, K. R.....	156, 166, 198, 209
Bauermeister, J. A.....	183
Baxter, R. K.....	118, 140, 141
Bergeron, D.....	168
Betjemann, R.....	70
Bettinger, G.....	171
Bieler, G.....	13, 63
Biemer, P.....	33, 41, 49, 123, 147, 177, 193, 194
Bierut, L. J.....	210
Biondo, G.....	210
Blazer, D. G.....	181
Bobashev, G. V.....	71, 141, 153
Bose, J.....	162, 167, 172, 179, 187, 194, 198, 199, 200, 201, 202, 204, 209, 217
Bowman, K.....	68, 128, 130, 133, 135, 142
Brantley, J.....	136
Bray, R. M.....	41
Brewer, R. D.....	117
Brittingham, A.....	57
Brodsky, M.....	21
Brown, G.....	123
Burke, B.....	42
Burns, R. M.....	197
Butler, D.....	95, 128, 142
Butler, J.....	180
Cajka, J.....	150
Camburn, D.....	77
Caspar, R.....	17, 18, 57, 58, 62, 63, 67, 73, 93, 103, 124, 128

Caulkins, J. P.	197
Cavazos-Rehg, P. A.	210
Center for Behavioral Health Statistics and Quality	183, 185, 197, 211, 213, 215
Chaloupka, F. J.	210
Chen, C. Y.	105
Chen, P.	71, 89, 164, 177, 193, 194, 203, 204
Chilcoat, H. D.	37, 210
Chromy, J.	68, 78, 81, 94, 97, 100, 108, 110, 113, 114, 125, 130, 138, 143, 150, 163, 165, 168, 169, 201
Cisin, I. H.	8, 11
Clark, C.	177, 180
Clarke, A.	110, 138
Cohen, M. P.	159
Colliver, J.	79, 170
Colpe, L. J.	107, 161, 163, 166, 173, 199
Copello, E.	147, 156
Couper, M. P.	43, 60
Couzens, A.	187
Cowan, C. D.	79
Cowell, A. J.	126
Cox, B. G.	19, 27
Cribb, D.	203
Crider, R. A.	9
Crits-Christoph, P.	155
Crump, C. J.	68
Cunningham, D.	118, 127, 133, 150, 166
Currihan, D.	178, 194, 205
Curry, R.	68
Curtin, G. M.	216
Cynamon, M.	45
Dai, L.	164, 203, 204
Davis, T.	94, 100, 156
Dean, E.	124, 171
Devore, J. W.	25
Díaz, A.	129
Dunn, R.	117
Durante, R. C.	24
Durell, T. M.	155
Enev, T.	149
Epstein, J.	34, 37, 54, 55, 59, 80, 107, 113, 156, 161
Everingham, S. S.	197

Eyerman, J.....	81, 95, 97, 108, 118, 119, 125, 128, 136, 137, 142, 158
Feder, M.....	143, 165
Fendrich, M.....	69, 81, 105
First, M. B.....	166
Fishburne, P. M.....	7
Flicker, L.....	127
Folsom, R. E., Jr.....	27, 35, 50, 72, 74, 90, 102, 106, 120, 153, 164, 206
Forman-Hoffman, V. L.....	188, 202
Forsyth, B. H.....	14, 20
Foster, M.....	204
Fournier, E.....	70
Fowler, F. J., Jr.....	82
Frechtel, P.....	116, 128, 136, 147, 187
George, B. J.....	26
Gerstein, D.....	39, 53, 61
Gfroerer, J.....	10, 13, 15, 21, 22, 23, 26, 34, 37, 39, 44, 48, 51, 52, 54, 55, 57, 59, 81, 82, 83, 94, 96, 97, 100, 104, 106, 107, 113, 114, 117, 132, 133, 136, 148, 161, 162, 163, 165, 166, 167, 169, 173, 179, 187, 188, 191, 198, 199, 201, 202, 209, 217
Giacoletti, K.....	97
Giles, W. H.....	117
Ginsberg, C.....	84
Gitelman, A. M.....	174
Glowacki, P.....	183
Gordek, H.....	75, 107, 164, 203
Graetz, K. A.....	14
Granger, R. A.....	110, 135, 138, 169, 171
Grau, E.....	90, 102, 103, 116, 120, 128
Graves, M. J.....	216
Groves, R. M.....	43, 60
Grucza, R. A.....	148, 210
Guess, L.....	48
Gustin, J.....	22
Haaga, J.....	29
Handley, W.....	156
Harrell, A. V.....	11, 52
Harrington, D.....	149
Harrison, E. R.....	29
Harrison, L.....	23, 84, 149
Hedden, S.....	179, 188, 199, 201
Heller, D.....	114, 167, 171
Henderson, L.....	209

Hennessy, K. H. ....	84
Hinsdale, M. ....	129
Hiripi, E. ....	107
Hirsch, E. ....	165, 169
Horm, J. ....	45
Hottinger, C. ....	137
Howes, M. J. ....	107
Hubbard, M.L. ....	14, 20, 24, 26, 123
Hughes, A. ....	13, 15, 22, 77, 79, 97, 108, 114, 125, 167, 168, 189, 200, 204, 206
Hunter, S. R. ....	130, 135, 143
Iannacchione, V. G. ....	150, 168, 171, 175, 190, 205
Johns, M. M. ....	183
Johnson, E. O. ....	131
Johnson, R. ....	39, 53, 61
Johnson, T. P. ....	69, 81, 88, 105
Jones, M. ....	162, 164
Jordan, B. K. ....	123, 156
Judkins, D. R. ....	50
Kann, L. ....	85, 187
Kapsak, K. A. ....	11
Karg, R. S. ....	156, 163, 166
Kay, W. ....	57
Kennet, J. ....	119, 127, 129, 132, 133, 134, 135, 137, 158, 162, 165, 169, 178
Kessler, R. C. ....	107
Kilmer, B. ....	197
Kilmer, G. ....	202
Kopstein, A. ....	52
Kott, P. ....	199, 201, 207
Krauss, M. J. ....	210
Kroutil, L. ....	41, 80, 98, 151, 155, 156, 170, 187, 194, 202
Kulka, R. A. ....	136
LaBrie, R. ....	70
Langer, M. ....	136
Lattimore, P. K. ....	217
Laufenberg, J. ....	164, 203
LeBaron, P. A. ....	171
Lepkowski, J. M. ....	85
Lessler, J. ....	14, 24, 25, 26, 51, 57, 58, 62, 63, 73
Liao, D. ....	199, 201
Liu, J. ....	35
Lomuto, N. ....	70

Lopez, M. ....	187
Lowery, A. ....	210
Lurigio, A. J. ....	145
Mamo, D. ....	126
Manderscheid, R. W. ....	107
Marano, K. M. ....	216
Marsden, M. E. ....	167
Martin, P. ....	128, 133, 143
Martin, S. S. ....	149
McAuliffe, W. E. ....	70, 116, 117
McCue, C. ....	137
McHenry, G. ....	177
McMichael, J. P. ....	150, 157, 168, 171, 172, 175, 190, 205
McNeeley, M. E. ....	108, 125, 136
Melnick, D. ....	54
Meyer, M. ....	110, 138
Miller, J. D. ....	12
Miller, J. W. ....	117
Miller, P. V. ....	54, 63
Mokdad, A. ....	117
Moore, A. ....	187, 207
Morton, K. B. ....	143, 150, 163, 165, 166, 168, 171, 172, 190, 200
Murphy, J. ....	118, 119, 127, 137, 141, 144, 158
Myers, L. ....	78, 98
Myers, S. ....	127
Naimi, T. S. ....	117
National Institute on Drug Abuse ....	36
Normand, S. L. ....	107
Novak, S. P. ....	141, 163, 173, 188
O'Rourke, D. ....	88
Odom, D. ....	81, 95, 97, 100, 108, 116, 125, 136
Office of Applied Studies ....	47, 86, 87, 99, 109, 138
Ogden, M. W. ....	216
Owens, L. ....	88
Packer, L. ....	68, 94, 100, 114
Painter, D. ....	110, 116, 134, 135, 136, 138, 140, 162, 177, 178, 194
Pan, J. J. ....	181
Pantula, J. ....	24, 27
Park, H. ....	144, 171
Parry, H. J. ....	8
Parsley, T. ....	29, 51

Paschall, M. J. ....	174
Peer Review Committee on National Household Survey on Drug Abuse.....	30
Pemberton, M.....	106, 188, 198, 202, 209, 217
Penne, M. ....	58, 63, 68, 71, 73, 89, 93, 106, 124, 139, 209, 217
Pepper, J. V. ....	89
Perez-Michael, A. M. ....	19
Peytcheva, E.....	178
Piper, L.....	104, 165, 169
Przybeck, T. R.....	148
Rasinski, K. A. ....	53, 61
Richards, T. ....	29
Ridenhour, J. L.....	157, 168, 171, 172, 175
Riggsbee, B. H. ....	165
Ringeisen, H.....	188
Ringwalt, C. L.....	174
Ryan, H. ....	191
Saavedra, L. M.....	188
Safir, A.....	144
Salinas, C. ....	129
Sathe, N.....	164, 189, 203, 204, 206
Schultz, L. ....	131
Schütz, C. G. ....	37
Schwerin, M.....	158
Scott, V. ....	187
Shah, B. V. ....	74
Shamblen, S. R.....	152
Shook-Sa, B. E.....	157, 168, 171, 175, 190, 200, 205
Singh, A. ....	71, 72, 74, 75, 78, 89, 90, 102, 107, 120
Smith, S.....	188
Snodgrass, J.....	103, 128, 129, 139, 165
Spagnola, K.....	163, 189, 206
Spiehler, V. ....	69
Spitznagel, E. L.....	210
Springer, J. F. ....	152
Steffey, D. M.....	217
Stivers, J.....	136
Stoddard, S.....	183
Stolzenberg, S. J.....	166, 168
Strashny, A.....	198, 209
Stringfellow, V. L. ....	82
Sudman, S. ....	69, 91

Sulsky, S. I. ....	216
Supple, A. J. ....	64
Swartz, J. A. ....	145, 158
Swauger, J. E. ....	216
Thoreson, R. ....	54
Thornberry, O. ....	45
Touarti, C. ....	180
Tourangeau, R. ....	57
Traccarella, M. A. ....	19
Trosclair, A. ....	191
Trunzo, D. ....	198, 209
Turner, C. F. ....	22, 25, 26
U.S. General Accounting Office ....	31, 64
Vaish, A. ....	74, 75, 189, 206
Van Brunt, D. L. ....	155
Van Landingham, C. ....	216
Virag, T. ....	42, 165
Volz, E. ....	183
Vorburger, M. ....	134, 151, 170
Walters, E. E. ....	107
Wang, K. ....	140, 141, 144, 147, 177, 193, 194, 207
Weimer, B. ....	209
Wheless, S. C. ....	41
Wiesen, C. ....	156
Williams, D. ....	177, 180
Williams, R. L. ....	13
Willson, D. ....	54
Wirtz, P. W. ....	11
Wislar, J. S. ....	69, 105
Witt, M. ....	19, 26, 27, 33, 41, 48, 49, 62
Woodward, A. ....	54, 103
Woody, G. E. ....	181
Wright, D. A. ....	37, 48, 52, 55, 65, 83, 104, 113, 120, 123, 128, 138, 141, 142, 153, 159
Wright, D. L. ....	64
Wu, L. T. ....	181
Wu, S. ....	95
Yang, C. ....	181
Yu, F. ....	165
Zaslavsky, A. M. ....	107
Zhang, Z. ....	65, 159
Zimmerman, M. A. ....	183

*This page intentionally left blank*

## Subject Index

- address-based sampling (ABS)..... 157, 168, 169, 171, 172, 175, 186, 190, 200, 205
- adolescent or adolescents..... 37, 58, 65, 66, 79, 82, 90, 91, 105, 141, 153, 156, 183, 184, 185, 187, 188, 202, 210
- Alcohol and Drug Services Study (ADSS)..... 136, 152
- American Community Survey (ACS)..... 196, 212, 215
- anchoring..... 15, 20
- any mental illness (AMI) ..... 189, 198, 199, 201
- Arrestee Drug Abuse Monitoring (ADAM) Program..... 217
- assessment or assessments ..... 19, 50, 64, 80, 85, 93, 95, 108, 113, 138, 143, 146, 152, 156, 161, 187, 198, 200, 203, 214
- audio computer-assisted self-interviewing (ACASI)..... 45, 57, 58, 62, 63, 67, 73, 80, 83, 86, 94, 95, 104, 106, 117, 124, 129, 139, 140, 166, 195, 212, 215, 216
- Behavioral Risk Factor Surveillance System (BRFSS)..... 77, 85, 88, 117, 174, 189, 202
- bias ..... 11, 15, 17, 18, 19, 20, 22, 25, 30, 33, 38, 39, 41, 45, 49, 51, 52, 53, 54, 56, 61, 62, 74, 75, 83, 86, 88, 96, 100, 105, 109, 113, 114, 115, 119, 120, 123, 124, 127, 131, 136, 137, 141, 142, 143, 147, 148, 155, 164, 167, 168, 172, 173, 177, 184, 193, 194, 198, 200, 203, 206, 209
- bilingual interviewers..... 130
- Brewer's method ..... 78, 113
- calibration ..... 71, 72, 73, 74, 75, 76, 78, 79, 89, 120, 121, 163, 164, 166, 167, 203, 204
- case management system (CMS)..... 178
- Check for Housing Units Missed (CHUM) ..... 157, 169, 173, 190, 191, 200
- cluster size variation ..... 78
- clustering..... 65, 71, 95, 143, 183
- cognitive interviewing or cognitive testing..... 14, 20, 27, 57, 62, 69, 86, 123, 130, 134, 135, 171
- cognitive process..... 20, 140
- cohort studies ..... 29, 109, 138, 207
- Composite International Diagnostic Interview (CIDI) ..... 80, 108, 159, 173
- Composite International Diagnostic Interview Short Form (CIDI-SF)..... 80, 108, 159, 173
- computer-assisted interviewing (CAI)..... 44, 51, 57, 62, 63, 67, 73, 81, 83, 86, 93, 94, 95, 96, 97, 98, 101, 102, 103, 104, 124, 129, 140, 184, 188, 196
- computer-assisted personal interviewing (CAPI)..... 45, 59, 86, 104, 106, 139, 195, 212, 215
- computer-assisted self-interviewing (CASI)..... 45, 46, 65, 108
- computerized screening ..... 81, 95, 96
- continuous quality improvement (CQI)..... 42
- coverage ..... 12, 32, 36, 44, 45, 52, 54, 55, 68, 74, 75, 79, 80, 82, 114, 134, 135, 143, 150, 151, 157, 169, 171, 172, 173, 175, 184, 188, 190, 191, 195, 196, 198, 199, 200, 201, 205, 209, 217, 218
- cross-sectional surveys..... 39, 53, 61, 62, 114, 131, 147, 167
- Current Population Survey (CPS)..... 127, 134, 196, 212, 215

data processing.....	101
design effect or design effects.....	13, 107, 113, 159
<i>Diagnostic and Statistical Manual of Mental Disorders (DSM)</i> .....	34, 108, 161, 166, 173, 181
distress scale.....	108
domain effects.....	13
dual-sample design.....	94
dwelling unit (DU) or dwelling units (DUs).....	25, 68, 71, 78, 89, 104, 111, 113, 127, 128, 143, 144, 157, 169, 172, 190, 193, 200, 201, 203, 212, 214
editing .....	19, 31, 47, 48, 58, 81, 82, 94, 95, 98, 99, 101, 102, 103, 151, 167
estimating equations.....	75, 105
estimation.....	12, 14, 33, 34, 35, 37, 38, 39, 41, 48, 49, 50, 54, 55, 56, 65, 66, 73, 74, 75, 78, 83, 84, 85, 86, 89, 91, 94, 101, 114, 120, 123, 142, 143, 159, 160, 162, 167, 170, 188, 189, 194, 198, 199, 200, 201, 206, 211, 213
evaluation.....	7, 20, 30, 31, 38, 43, 78, 79, 82, 85, 94, 116, 124, 132, 133, 151, 157, 172, 189, 195, 198, 209, 216
exponential decay.....	62
extreme values .....	72, 73, 89
extreme weight or extreme weights .....	79, 89, 144
face-to-face survey or face-to-face surveys .....	9, 81, 124, 139, 147
falsification .....	118, 119, 137, 180
field enumeration (FE).....	150, 151, 157, 168, 169, 172, 190, 200, 205
field experiment .....	27, 59, 62, 63, 80, 86, 87
field interviewer (FI) or field interviewers (FIs).....	43, 69, 81, 96, 98, 100, 108, 110, 119, 134, 135, 136, 137, 140, 144, 158, 165, 169, 180, 200, 201, 207, 212, 214
follow-up survey.....	17, 21
generalized exponential model (GEM) and/or modeling.....	71, 72, 73, 75, 89, 120, 164, 193
generalized variance functions (GVF).....	13, 107
geographic information system (GIS).....	68
group quarters (GQs) .....	145, 165, 173, 180, 191, 201
Hajek-type modification .....	79
half-open interval (HOI) .....	151, 169, 172, 190, 200
hand-held computer .....	104
hard-to-reach .....	12, 32
hot-deck.....	91, 102, 116, 120
household roster .....	113
housing unit (HU) .....	30, 60, 61, 128, 150, 205
Hui-Walter method .....	33, 41, 49, 50
hybrid designs.....	186
hybrid frame or hybrid frames .....	169, 172, 173, 191, 201
hybrid sampling .....	190, 191, 200, 201

imputation .....	12, 47, 90, 91, 94, 95, 99, 101, 102, 103, 116, 120, 147, 148, 167, 184, 187, 196, 212, 214, 215
incentive or incentives .....	7, 17, 18, 19, 69, 87, 99, 100, 109, 110, 111, 120, 128, 133, 134, 135, 136, 138, 139, 142, 149, 158, 165, 167, 170
instrument design .....	132, 133
interview mode .....	46, 58
interviewer experience .....	94, 97, 108, 109, 125, 126, 186, 207, 208
intracluster correlation .....	78
item count .....	123
item nonresponse .....	7, 27, 47, 88, 124, 125, 128, 196, 212
Kessler-6 (K6) .....	108, 145, 146, 159, 161, 163, 173, 179, 201
Kessler-10 (K10) .....	108, 173
level of effort (LOE) .....	147, 177, 193, 203
linear models .....	65, 105, 107
log-linear model .....	13
longitudinal study .....	114
major depressive episode (MDE) .....	80, 161, 189, 201
mapping .....	68, 69
mean squared error (MSE) .....	35, 49, 120, 207
measurement error .....	20, 51, 124, 131
Medical Expenditure Panel Survey (MEPS) .....	189, 202
Mental Health Surveillance Study (MHSS) .....	163, 166, 179, 198, 200, 201
metadata .....	118, 119, 137
micro agglomeration, substitution, subsampling, and calibration (MASSC) .....	121
misclassification .....	33, 41, 49, 163, 200
Monitoring the Future (MTF) .....	52, 79, 80, 81, 82, 84, 85, 89, 91, 115, 183, 188, 210
multicultural .....	129, 130
multilevel logistic model .....	51
multiplicities .....	42, 78, 79
multiplicity factors .....	89, 91
multistage area probability sample .....	65, 104, 139
National Comorbidity Survey Replication (NCS-R) .....	161, 189
National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) .....	148, 189
National Health Interview Survey (NHIS) .....	46, 189, 191, 195, 202, 212, 215, 216
National Survey of Family Growth (NSFG) .....	215, 216
National Survey of Substance Abuse Treatment Services (N-SSATS) .....	198, 199, 209
nearest neighbor hot deck (NNHD) .....	91, 116, 120
nominative technique .....	12
noncoverage .....	13, 23, 68, 86
nonresponse bias .....	17, 18, 19, 52, 74, 86, 100, 119, 141, 142, 147, 148, 184, 194, 203
official records .....	52

older adults.....	30, 51, 103, 104, 106, 158, 170
overreport or overreporting.....	8, 54, 134, 150, 168, 184, 188
pair data analysis.....	79, 89, 91
paper-and-pencil interviewing (PAPI).....	57, 62, 63, 65, 67, 80, 81, 83, 86, 93, 94, 95, 96, 97, 98, 101, 102, 103, 104, 106, 129, 140
personal interview or personal interviews .....	15, 39, 42, 43
person-level nonresponse (PRNR).....	203
Pittsburgh Adolescent Alcohol Research Center's Structured Clinical Interview (PAARC-SCID).....	156
Population Assessment of Tobacco and Health (PATH) survey .....	216
population estimates.....	100, 108, 128, 137, 139
poststratification.....	72, 74, 75, 144, 164, 203, 204
predictive mean.....	90, 91, 102, 116, 120, 187
predictive mean neighborhood (PMN) .....	90, 91, 102, 116, 120, 187
privacy.....	10, 11, 12, 21, 25, 26, 37, 46, 52, 53, 57, 58, 63, 64, 73, 80, 87, 94, 135, 140, 149, 150, 158, 168, 184, 188
quality control.....	109, 110, 138
question wording or question wordings.....	20, 25, 26, 91, 101, 130, 135, 184, 188, 196, 202
questionnaire changes.....	26, 47, 60, 67, 101, 103, 109, 129, 132, 133, 138, 140, 149, 167, 171, 186, 194, 195, 196, 197, 211, 213
raking.....	73
random-digit dialing (RDD) .....	155
recall decay .....	53, 114
record of calls (ROC).....	110, 119, 137, 138, 139, 193
reliability.....	49, 70, 117, 135, 143, 165, 169, 170
repeated measures .....	33, 41, 49
respondent characteristics .....	65, 88, 111, 125, 163
respondent errors.....	59
response bias or response biases.....	20, 33, 45, 49, 61, 62, 100, 105, 113, 136, 142
response error or response errors.....	53, 61
response propensity.....	51, 110, 142, 147, 177, 193
retrospective report .....	10, 21, 22, 53, 100, 109, 114, 115, 131, 138
sampling error or sampling errors.....	47, 54, 115, 167
sampling frame.....	54, 150, 157, 168, 169, 171, 172, 190, 191, 200, 201, 205
sandwich variance.....	75
screening or screenings.....	18, 43, 53, 57, 58, 62, 68, 81, 83, 86, 95, 96, 104, 107, 108, 109, 111, 113, 119, 126, 136, 137, 139, 143, 144, 145, 146, 159, 161, 163, 171, 172, 173, 178, 180, 186, 193, 211, 213, 215
selection probabilities .....	79, 81, 89
self-administered questionnaires.....	10, 14, 24, 45, 57, 58, 63, 64, 65, 67, 94, 108, 124
self-report or self-reported .....	9, 10, 11, 12, 23, 25, 29, 31, 33, 41, 49, 51, 52, 53, 55, 61, 62, 65, 69, 80, 90, 100, 108, 123, 124, 125, 126, 149, 150, 167, 168, 195, 202, 208, 217, 218

sensitive questions .....	15, 16, 46, 65, 105, 139, 171
sensitive topics .....	14, 37, 44, 45, 46, 140
serious mental illness (SMI) .....	107, 109, 145, 158, 159, 161, 163, 166, 173, 179, 189, 198, 199, 201
serious psychological distress (SPD) .....	162, 189
Sheehan Disability Scale (SDS) .....	163
small area estimation (SAE) .....	35, 50, 86, 189, 206
social desirability .....	54
Spanish or Spanish language .....	129, 130, 178, 179, 214, 215, 216
Structured Clinical Interview for DSM-IV (SCID-IV) .....	156, 163
substance use disorder (SUD) or substance use disorders .....	113, 116, 145, 146, 149, 156, 161, 181, 189
survey measurement .....	17, 19, 20, 22, 24, 25, 26, 27, 89
survey weights .....	74
synthetic estimates .....	206, 207
Taylor linearization .....	75, 143
technical advisory group .....	179
telephone survey or telephone surveys .....	13, 14, 15, 23, 77, 86, 155
telescoping .....	53, 114, 131
test-retest .....	33, 41, 49, 61, 135
text to speech (TTS) .....	215, 216
time series .....	9, 94
timing data .....	118, 140, 171, 195
Tobacco Use Supplement to the Current Population Survey (TUS-CPS) .....	216
translation .....	129, 130
Treatment Episode Data Set (TEDS) .....	198, 199, 209
treatment need or treatment needs .....	34, 59, 60, 64, 106, 113, 116, 117, 119, 209, 217
trend data .....	114, 196
trend or trends .....	9, 10, 12, 21, 29, 30, 31, 38, 39, 42, 53, 61, 62, 82, 83, 84, 85, 87, 89, 90, 94, 99, 104, 109, 114, 115, 138, 139, 157, 162, 167, 171, 184, 186, 188, 189, 192, 195, 196, 197, 200, 201, 210, 211, 213
undercoverage .....	38, 56, 68, 114, 116, 173, 175, 186, 191, 199, 218
underreport or underreporting .....	8, 9, 10, 21, 32, 38, 45, 51, 52, 54, 55, 56, 62, 83, 105, 114, 116, 124, 134, 150, 168, 170, 193
unequal weighting effect (UWE) .....	72, 78, 89, 143
Uniform Reporting System (URS) .....	189
United States Postal Service Computerized Delivery Sequence (USPS CDS) .....	172, 201, 205
usability testing .....	67, 171
validation or validity .....	8, 9, 10, 11, 12, 21, 22, 33, 41, 44, 49, 51, 52, 53, 54, 55, 61, 69, 70, 76, 80, 84, 91, 114, 117, 123, 149, 150, 156, 167, 168, 169, 174, 190
variance models or variance modeling .....	78, 130

verification .....	118, 140, 180
weight calibration.....	71, 72, 75, 89, 120, 203, 204
weighting or weighted.....	14, 31, 38, 48, 56, 68, 72, 74, 78, 79, 80, 89, 91, 99, 107, 109, 110, 116, 138, 142, 143, 144, 147, 155, 163, 164, 165, 167, 168, 177, 184, 202, 203, 204, 206, 216
World Health Organization Disability Assessment Schedule (WHODAS) .....	108, 163, 173, 201
Youth Risk Behavior Surveillance System (YRBSS) .....	91, 183, 188, 210
Youth Risk Behavior Survey (YRBS) .....	46, 79, 80, 81, 82, 84, 85, 174, 183, 188, 210